



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 12:21:36 ; Search time 90 Seconds

(without alignments)  
2054.627 Million cell updates/sec

Title: US-10-016-768A-10

Perfect score: 6030  
Sequence: 1 MHISYSYISLIERVAECWGR.....ILHEKLAQIKAEVQDADQL 1165

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6030	100.0	1165	23	AAE24372
2	5882	97.5	1140	22	ABR71145
3	4839	80.2	1221	23	AAE24373
4	314.5	5.2	1937	22	ABBS5895
5	314.5	5.2	2151	22	ABBS6086
6	311.5	5.0	2112	22	ABBS6403
7	299.5	4.9	2175	22	ABBS6598
8	296.5	4.9	1161	22	ABR70667
9	294	4.9	2441	22	ABBS2231

10	293	4.9	1436	22	ABBS1959	Drosophila melanog
11	287	4.8	1369	22	ABBS6040	Drosophila melanog
12	285.5	4.7	1409	22	ABBS1328	Drosophila melanog
13	285	4.7	1703	22	ABBS6223	Drosophila melanog
14	283	4.7	1416	22	ABBS5499	Drosophila melanog
15	282.5	4.7	1741	22	ABBS6091	Drosophila melanog
16	281.5	4.7	1198	22	ABBS6371	Drosophila melanog
17	281	4.7	2280	22	ABBS1650	Drosophila melanog
18	279.5	4.6	1186	22	ABBS1516	Drosophila melanog
19	279	4.6	2110	22	ABBS8077	Drosophila melanog
20	279	4.6	2150	23	AAO22566	Wooden leg (WOL) g
21	278.5	4.6	1665	22	ABBS64010	Drosophila melanog
22	278	4.6	53	23	AAE24370	Fruit fly E93 prog
23	276	4.6	1078	22	ABBS58620	Drosophila melanog
24	274.5	4.6	3502	22	ABBS58382	Drosophila melanog
25	274	4.5	2344	22	AAU37120	Staphylococcus aur
26	273.5	4.5	1164	22	ABBS7802	Drosophila melanog
27	272.5	4.5	1246	22	ABBS1547	Drosophila melanog
28	272.5	4.5	1669	22	ABBS64003	Drosophila melanog
29	271	4.5	2309	22	ABBS6232	Drosophila melanog
30	270.5	4.5	1217	24	ABBS60377	Bombyx mori serici
31	270	4.5	1467	23	ABBS97605	Novel human protei
32	268.5	4.5	1368	22	ABBS60262	Drosophila melanog
33	267.5	4.4	1920	22	ABBS4441	Drosophila melanog
34	266	4.4	1028	22	ABBS2708	Drosophila melanog
35	266	4.4	2283	22	ABBS6876	Staphylococcus epi
36	263.5	4.4	2261	24	ABJ18914	Pathogen specific
37	262	4.3	1911	22	ABBS1038	Drosophila melanog
38	261.5	4.3	1326	22	ABBS1156	Drosophila melanog
39	261	4.3	2451	22	ABBS1574	Drosophila melanog
40	259.5	4.3	1545	22	ABBS5577	Drosophila melanog
41	257	4.3	738	22	ABBS5805	Drosophila melanog
42	257	4.3	1001	22	ABBS6527	Drosophila melanog
43	257	4.3	1805	22	ABBS5262	Drosophila melanog
44	256.5	4.3	1059	22	ABBS5792	Drosophila melanog
45	256.5	4.3	1077	22	ABBS1539	Drosophila melanog

#### ALIGNMENTS

RESULT 1	
AAE24372	
ID	AAE24372 standard; Protein: 1165 AA.
AC	AAE24372;
XX	
DT	04-OCT-2002 (first entry)
XX	
DE	Fruit fly E93 programmed cell death modulating protein #1.
XX	
KW	Fruit fly; programmed cell death modulating protein; adenocarcinoma;
KW	cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW	neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;
KW	Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;
KW	aplastic anaemia; ischaemic injury; myocardial infarction; stroke;
KW	reperfusion injury; toxin-induced disease; genetic immunodeficiency;
KW	vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;
KW	vaccine; neurotropic; vasotrophic; immunostimulant; cerebroprotective;
KW	cardiant; cancer; E93 protein.
OS	
XX	
PN	Drosophila melanogaster.
XX	
PD	WO20024882-A2.
XX	
PF	02-MAY-2002.
XX	
PR	29-OCT-2001; 2001WO-US48053.
XX	
PR	27-OCT-2000; 2000US-243865P.
XX	
PA	(UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.
XX	

PI Baehrecke EH;  
 XX WPI: 2002-479717/51.  
 DR N-PSDB; AAD39237.  
 XX Novel programmed cell death modulating proteins, useful for treating or  
 PT preventing disorders associated with abnormal cell proliferation and  
 PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
 PT infarction -  
 XX  
 PS Claim 9; Page 65-71; 88pp; English.  
 XX  
 CC The present invention relates to novel programmed cell death modulating  
 CC proteins and polynucleotides encoding such proteins. Sequences of the  
 CC invention are useful to screen potential cellular apoptosis inhibiting  
 CC compounds to determine their use as therapeutic agents for treatment of  
 CC diseases associated with increased programmed cell death. They are also  
 CC useful for treating or preventing disorders associated with decrease in  
 CC apoptosis. Programmed cell death modulating sequences are useful for  
 CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,  
 CC melanoma, myeloma. Inhibition of the activity of the sequences of the  
 CC invention are useful for treating disorders associated with increase  
 CC in cell death or apoptosis such as acquired immunodeficiency syndrome  
 CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis  
 CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischemic  
 CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),  
 CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced  
 CC diseases and other infectious or genetic immunodeficiencies. Sequences  
 CC of the invention are used as vaccines and in gene therapy. The present  
 CC sequence is fruit fly E93 programmed cell death modulating protein.  
 XX  
 SO Sequence 1165 AA:  
 Query Match 100.0%; Score 6030; DB 23; Length 1165;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 541 QQQQQQVAAVRRRLPKSETPETNSSLPNDASEDPLTKIPSFVSGPASSSSLSPGGLV 600  
 DB 541 QQQQQQVAAVRRRLPKSETPETNSSLPNDASEDPLTKIPSFVSGPASSSSLSPGGLV 600  
 QY 601 GGHHPPLNNNSLSISNNNSHSSNSHRNGSNRSPHSASPMIAAAVAGGYAGNSLLTSS 660  
 DB 601 GGHHPPLNNNSLSISNNNSHSSNSHRNGSNRSPHSASPMIAAAVAGGYAGNSLLTSS 660  
 QY 661 SSSIQKMAASVIOQINQESQESLRNGVSDCSNNGSSSLGKPPSIVAKIIGTD 720  
 DB 661 SSSIQKMAASVIOQINQESQESLRNGVSDCSNNGSSSLGKPPSIVAKIIGTD 720  
 QY 721 TSFSGAPNLLSQOHSHAHLLTHQOQQOQSLAQALGKTRPKKGRNDRDRLVEAVK 780  
 DB 721 TSFSGAPNLLSQOHSHAHLLTHQOQQOQSLAQALGKTRPKKGRNDRDRLVEAVK 780  
 QY 781 AVQGENSVHRAGSYGVPHSTLEYKVERHLMRPRRREPPODVLGLTGPAKQLDK 840  
 DB 781 AVQGENSVHRAGSYGVPHSTLEYKVERHLMRPRRREPPODVLGLTGPAKQLDK 840  
 QY 841 LKAGPHGSKLSNALKNNQNNQAAAAAATPNGIKLPLFAGPOLSFQPMF 900  
 DB 841 LKAGPHGSKLSNALKNNQNNQAAAAAATPNGIKLPLFAGPOLSFQPMF 900  
 QY 901 WPTNATNAYGLDENRITTEAMRNPOASHHGLMKSADWENVYDGIIRKTLQASENGS 960  
 DB 901 WPTNATNAYGLDENRITTEAMRNPOASHHGLMKSADWENVYDGIIRKTLQASENGS 960  
 QY 961 AAGNGSNGNGHGHGHGHALLDLVLKTPPLPTNRHNDYATTCSSASGESVRS 1020  
 DB 961 AAGNGSNGNGHGHGHGHALLDLVLKTPPLPTNRHNDYATTCSSASGESVRS 1020  
 QY 1021 GSPWGNVADIKREKLSADSGSSSDEHSASHINNNSDLAHNKKSGGGGGGNGCQTNG 1080  
 DB 1021 GSPWGNVADIKREKLSADSGSSSDEHSASHINNNSDLAHNKKSGGGGGGNGCQTNG 1080  
 QY 1081 NGRSSRMTSRDSDSTDSFSGENGGQONKMDLNGSSSSSHICESEPAATGHHSPG 1140  
 DB 1081 NGRSSRMTSRDSDSTDSFSGENGGQONKMDLNGSSSSSHICESEPAATGHHSPG 1140  
 QY 1141 HTTSTLHEKLAQIKAEVDQADQL 1165  
 DB 1141 HTTSTLHEKLAQIKAEVDQADQL 1165

RESULT 2  
 ABB71145  
 ID ABB71145 standard; Protein; 1140 AA.  
 XX  
 AC ABB71145;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 40227.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 OS  
 XX Drosophila melanogaster.  
 OS  
 XX  
 PN WC00171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PA  
 XX Venter JC, Adams M, Li FMD, Myers EW,  
 XX



DR WPI: 2001-65660/75.  
 DR N-PSDB; ABL15248.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS  
 PS  
 PS Disclosure; SEQ ID NO 40227; 21np + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL130511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AAB57737-AB12072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC  
 CC Sequence 1140 AA;  
 XX  
 XX  
 Query Match 97.5%; Score 5882; DB 22; Length 1140;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 1135; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Oy 18 MGRQWKHYODKLTCTSHNIEEQPIAAGSEDEPSQYNSKESKISQNPCHKTEHRL 77  
 Db 1 MGRQWKHYODKLTCTSHNIEEQPIAAGSEDEPSQYNSKESKISQNPCHKTEHRL 60  
 Oy 78 EQHNGSOLLEBEDSENNQTSHTSRTPTPGATSPSPPEPIDMRPSAKCNFCVNGRL 137  
 Db 61 EQHNGSOLLEBEDSENNQTSHTSRTPTPGATSPSPPEPIDMRPSAKCNFCVNGRL 120  
 Oy 138 TVNAGKLVASAAATATSSSTNSHIOHSDSDNSASLPHNIISSSSSSNNNSGNRAH 197  
 Db 121 TVNAGKLVASAAATATSSSTNSHIOHSDSDNSASLPHNIISSSSSSNNNSGNRAH 180  
 Oy 198 IAAASARTPAAATPANSLELYKLTORAAKMTSDMSAAQLAQSLLADFLINLSAQ 257  
 Db 181 IAAASARTPAAATPANSLELYKLTORAAKMTSDMSAAQLAQSLLADFLINLSAQ 240  
 Oy 258 QQQQQQQQIASAAPTSTSEVSAATSPALKDTPSPSVAPRLDLSKSPNSISGDVASV 317  
 Db 241 QQQQQQQQIASAAPTSTSEVSAATSPALKDTPSPSVAPRLDLSKSPNSISGDVASV 300  
 Oy 318 RACATPTSGRRAYSEEDLSRALQDVANKLDARKSASQHEORSILDNRLFMMGHNDQ 377  
 Db 301 RACATPTSGRRAYSEEDLSRALQDVANKLDARKSASQHEORSILDNRLFMMGHNDQ 360  
 Oy 378 ODHGDDELSDNDAAEAVDSNAATPVYPAEFARQLKLSLSHNSDGLGEDVDKRS 437  
 Db 361 ODHGDDELSDNDAAEAVDSNAATPVYPAEFARQLKLSLSHNSDGLGEDVDKRS 420  
 Oy 438 KMGHRPACGNASANGAPASIPLDANVLLHTMLAAGIGAMPKLDCTQVDFIKGLVA 497  
 Db 421 KMGHRPACGNASANGAPASIPLDANVLLHTMLAAGIGAMPKLDCTQVDFIKGLVA 480  
 Oy 498 NSGIMNEGLNLTLASQENSNGNASLLQQQOHOQHQQOHHQQOQQOHHVAAVYHRLPK 557  
 Db 481 NSGIMNEGLNLTLASQENSNGNASLLQQQOHOQHQQOHHQQOQQOHHVAAVYHRLPK 540  
 Oy 558 SETPTNSLDPNDASEDPILKIPSFKVSGPASSSSLSPGLVGHHHPPLNNNSLSISN 617  
 Db 541 SETPTNSLDPNDASEDPILKIPSFKVSGPASSSSLSPGLVGHHHPPLNNNSLSISN 600  
 Oy 618 NSNHSNHRNGSNNSPHSASPMLAAVAAGYSAGNSLTLSSSSSIKMAASNTORQIN 677  
 Db 601 NSNHSNHRNGSNNSPHSASPMLAAVAAGYSAGNSLTLSSSSSIKMAASNTORQIN 660  
 Oy 678 EQSGGESLRNGNVSDCSSNNGSSSLGYKKPSISVAKIIGTDTSRFGASPNLLSQOHS 737

Db 661 EQSGGESLRNGNVSDCSSNNGSSSLGYKKPSISVAKIIGTDTSRFGASPNLLSQOHS 720  
 Oy 738 AHHLTHQOOOQOOLSAQOALGKGTBPKRGKYNVYRDSIYEAVKAVQREMSVHRAGSYG 797  
 Db 721 AHHLTHQOOOQOOLSAQOALGKGTBPKRGKYNVYRDSIYEAVKAVQREMSVHRAGSYG 780  
 Oy 798 VPHSTLEYKVERHLMRPRKREPKQPDVLVGLTGPANKLQDLKAKGPHGSKLSNALKN 857  
 Db 781 VPHSTLEYKVERHLMRPRKREPKQPDVLVGLTGPANKLQDLKAKGPHGSKLSNALKN 840  
 Oy 858 QNNQAAAAAATAATNGKLPLFEAGPOLSFQPMFMPQTNATNAYGIDFNRI 917  
 Db 841 QNNQAAAAAATAATNGKLPLFEAGPOLSFQPMFMPQTNATNAYGIDFNRI 900  
 Oy 918 TEARNPNQASVHNGHGMASQOMVENVDGILRTKLQASEGNSAAGNSNSNGHGHG 977  
 Db 901 TEARNPNQASVHNGHGMASQOMVENVDGILRTKLQASEGNSAAGNSNSNGHGHG 960  
 Oy 978 HGHGALLDQLLVKKTPLPTNHRNDYAATCSSASGESVVRSGSPMGNAVADIKERLSA 1037  
 Db 961 HGHGALLDQLLVKKTPLPTNHRNDYAATCSSASGESVVRSGSPMGNAVADIKERLSA 1020  
 Oy 1038 DSGGSSDEHSHASHIINNNSDLAHNKKSGGGGGGNGQTNNGRSSRMTSRDDETD 1097  
 Db 1021 DSGGSSDEHSHASHIINNNSDLAHNKKSGGGGGGNGQTNNGRSSRMTSRDDETD 1080  
 Oy 1098 SSFSGENGGQONIKMDLNGSSSSSHIKESEAATHHSPGHHTTSILHEKLAQICAE 1157  
 Db 1081 SLSKSGESGGQONIKMDLNGSSSSSHIKESEAATHHSPGHHTTSILHEKLAQICAK 1140  
 RESULT 3  
 AAE24373  
 ID AAE24373 standard; Protein; 1221 AA.  
 XX  
 AC AAE24373;  
 XX  
 DT 04-OCT-2002 (first entry)  
 XX  
 DE Fruit fly E93 programmed cell death modulating protein #2.  
 KW Fruit fly; programmed cell death modulating protein; adenocarcinoma;  
 KW cellular apoptosis; leukemia; acquired immunodeficiency syndrome; AIDS;  
 KW neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;  
 KW Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;  
 KW aplastic anaemia; ischaemic injury; myocardial infarction; stroke;  
 KW reperfusion injury; toxin-induced disease; genetic immunodeficiency;  
 KW vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;  
 KW myeloma; nocrotropic; vasotrophic; immunostimulant; cerebroprotective;  
 KW cardiact; cancer; E93 protein.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200234882-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-US48053.  
 XX  
 PR 27-OCT-2000; 2000US-243865P.  
 XX  
 PA (UTMA-) UNIV MARYLAND BIOTECHNOLOGY INST.  
 XX  
 PT Baehrecke EH;  
 XX  
 DR WPI: 2002-479717/51.  
 DR N-PSDB; AAD39238.  
 XX  
 PT Novel programmed cell death modulating proteins, useful for treating or  
 PT preventing disorders associated with abnormal cell proliferation and  
 PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
 PT infarction -  
 XX

PS Disclosure: Page 77-82; 88pp; English.

CC The present invention relates to novel programmed cell death modulating  
 CC proteins and polynucleotides encoding such proteins. Sequences of the  
 CC invention are useful to screen potential cellular apoptosis inhibiting  
 CC compounds to determine their use as therapeutic agents for treatment of  
 CC diseases associated with increased programmed cell death. They are also  
 CC useful for treating or preventing disorders associated with decrease in  
 CC apoptosis. Programmed cell death modulating sequences are useful for  
 CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,  
 CC melanoma, myeloma. Inhibition of the activity of the sequences of the  
 CC invention are useful for treating disorders associated with increase  
 CC in cell death or apoptosis such as acquired immunodeficiency syndrome  
 CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis  
 CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischemic  
 CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),  
 CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced  
 CC diseases and other infectious or genetic immunodeficiencies. Sequences  
 CC of the invention are used as vaccines and in gene therapy. The present  
 CC sequence is fruit fly B33 programmed cell death modulating protein.

XX Sequence 1221 AA;

Query Match 80.2%; Score 4839; DB 23; Length 1221;  
 Best Local Similarity 84.0%; Pred. No. 0;  
 Matches 972; Conservative 25; Mismatches 90; Indels 70; Gaps 9;

QY 1 MHISSEIETSERVAECCGRGROMKHYODKLTCSHNTIEEQPIAAGEDEPSQVNHSSK 60  
 DB 1 MHISSEIETSERVAECCGRGROMKHYODKLTCSHNTIEEQPIAAGEDEPSQVNHSSK 60  
 QY 61 EISQSNPNHCKTEHRLTEQHNHSGQLLEEDENNQTHSDSRTPTPATSTPSPPPPI 120  
 DB 61 EISQSNPNHCKTEHRLTEQHNHSGQLLEEDENNQTHSDSRTPTPATSTPSPPPPI 120  
 QY 121 DWRPSAKNFCVNGRLLTVNAQKLVASAAATATSSSTNSHIIHQHDSNSASLPHHI 180  
 DB 121 DWRPSAKNFCVNGRLLTVNAQKLVASAAATATSSSTNSHIIHQHDSNSASLPHHI 180  
 QY 181 SSSSSNNNSNGNRARHIAAASARATPAATPANSLELYKLTQPAAKMTSDSAAOLA 240  
 DB 181 SSSSSNNNSNGNRARHIAAASARATPAATPANSLELYKLTQPAAKMTSDSAAOLA 240  
 QY 241 QPSLLADNLNLSLAQO 300  
 DB 241 QPSLLADNLNLSLAQO 300  
 QY 301 SSKPSNNSISGDKSVACATPTPSGRAYSEEDLSRALQDVANKLDARKSAQHQHEQ 360  
 DB 301 SSKPSNNSISGDKSVACATPTPSGRAYSEEDLSRALQDVANKLDARKSAQHQHEQ 360  
 QY 361 RSLIDNRLFKKHHQOEDHDDGDELEDNDAAEAVDVSNAATPPVPAEFARQRLKLSHL 420  
 DB 361 RSLIDNRLFKKHHQOEDHDDGDELEDNDAAEAVDVSNAATPPVPAEFARQRLKLSHL 420  
 QY 421 SEHNSDGEVDVDRGSPKMGRRPACGNASANGAPASIPLDANVLLHTLMLAAGIGAMPK 480  
 DB 421 SEHNSDGEVDVDRGSPKMGRRPACGNASANGAPASIPLDANVLLHTLMLAAGIGAMPK 480  
 QY 481 LDETQVDFIKGLLVANSGLIMNGLNLTLASQENSNGNASLLQOQOQHQQHQHQHQ 540  
 DB 481 LDETQVDFIKGLLVANSGLIMNGLNLTLASQENSNGNASLLQOQOQHQQHQHQHQ 540  
 QY 541 QOQOQOQAAVRRHRLPKSETPEPTNSLDPNDASEPILKIPFKVSGPSSSSLSPPGLV 600  
 DB 541 QOQOQOQAAVRRHRLPKSETPEPTNSLDPNDASEPILKIPFKVSGPSSSSLSPPGLV 600  
 QY 601 GGHHPPLNNNSLSLSSNSNSHNRGNSNPSASPMIAAAGGYSAGNSLTS 660  
 DB 601 GGHHPPLNNNSLSLSSNSNSHNRGNSNPSASPMIAAAGGYSAGNSLTS 660  
 QY 661 SSIQKMAASNIQROINEQSGESLRNGVSDCSNNGSSSLGYKKPSIVAKIIGTGD 720  
 DB 661 SSIQKMAASNIQROINEQSGESLRNGVSDCSNNGSSSLGYKKPSIVAKIIGTGD 720

DB 661 SSIQKMAASNIQROINEQSGESLRNGVSDCSNNGSSSLGYKKPSIVAKIIGTGD 720  
 QY 721 TSREGAPNLSQOHHAAHHTHOQO 775  
 DB 721 TSREGAPNLSQOHHAAHHTHOQO 780  
 QY 776 -VEAVKAVQOREMNVHAGSYGV-----PHSTLEYKVKERHLMRPKREPKQPDIVG 828  
 DB 776 -VEAVKAVQOREMNVHAGSYGV-----PHSTLEYKVKERHLMRPKREPKQPDIVG 828  
 QY 829 LT--GPANKIQLDKXKAGPH-----GGSKLSSNALKNQNNQAAAAAATAATP 878  
 DB 829 LT--GPANKIQLDKXKAGPH-----GGSKLSSNALKNQNNQAAAAAATAATP 878  
 QY 839 QTEGGTTWMLQAEQCPQEPKQSGCGGCGGSSSGRCHQAPRPAAPRGSGTGVLISA-- 896  
 DB 839 QTEGGTTWMLQAEQCPQEPKQSGCGGCGGSSSGRCHQAPRPAAPRGSGTGVLISA-- 896  
 QY 879 NGLKLPLFEAGPQALSTQPMNFWPQTATNAYGIDFPRKITPAMNPPQASNHG----- 931  
 DB 879 NGLKLPLFEAGPQALSTQPMNFWPQTATNAYGIDFPRKITPAMNPPQASNHG----- 931  
 QY 932 -----LMKSAQDMEVENVYDGIIRKTLQASREGNSAAGNSGNSGNGHGHGHALLD 986  
 DB 932 -----LMKSAQDMEVENVYDGIIRKTLQASREGNSAAGNSGNSGNGHGHGHALLD 986  
 QY 926 LQSPRLKMSAQDMEVGERLPMHMHQEDAAAGEQNSAAGNSGNSGNGHGHGHALLD 985  
 DB 926 LQSPRLKMSAQDMEVGERLPMHMHQEDAAAGEQNSAAGNSGNSGNGHGHGHALLD 985  
 QY 987 QLVKKTPLPPTNHRNDYAAATCSSASGESYKRSQSGPMGNVADIKRRLSADSGSSDEE 1046  
 DB 987 QLVKKTPLPPTNHRNDYAAATCSSASGESYKRSQSGPMGNVADIKRRLSADSGSSDEE 1046  
 QY 1047 HSASHINNNSDLAHNNKSGCGGCGGNGOCTNGRSSRTPSRDDETTAASSFKSGENG 1106  
 DB 1047 HSASHINNNSDLAHNNKSGCGGCGGNGOCTNGRSSRTPSRDDETTAASSFKSGENG 1106  
 QY 1046 HSASHINNNSDLAHNNKSGCGGCGGNGOCTNGRSSRTPSRDDETTAASSFKSGENG 1104  
 DB 1046 HSASHINNNSDLAHNNKSGCGGCGGNGOCTNGRSSRTPSRDDETTAASSFKSGENG 1104  
 QY 1107 GQONHKMMDLNGSGSSS 1123  
 DB 1107 GQONHKMMDLNGSGRAA 1121

RESULT 4  
 ABB58985  
 ID ABB58985 standard; Protein; 1937 AA.  
 AC ABB58985;  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 3747.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 OS Drosophila melanogaster.  
 PN WO200171042-A2.  
 XX 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US09231.  
 PF 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 PA Venter JC, Adams M, Li FWD, Myers EW;  
 XX WPI; 2001-6556860/75.  
 DR N-PSDB; ABL03088.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Disclosure; SEQ ID NO 3747; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent











[illegible]

ID	AB062231	standard; Protein; 2441 AA.
XX	AB062231;	
XX	26-MAR-2002	(first entry)
XX	Drosophila melanogaster polypeptide SEQ ID NO 13485.	
XX	Drosophila; developmental biology; cell signalling; insecticide;	
XX	pharmaceutical.	
XX	Drosophila melanogaster.	
XX	WO200171042-A2.	
XX	27-SEP-2001.	
XX	23-MAR-2001; 2001WO-US09231.	
XX	23-MAR-2000; 2000US-191637P.	
XX	11-JUL-2000; 2000US-0614150.	
XX	(PEKE ) PE CORP NY.	
XX	Venter JC, Adams M, Li PWD, Myers EW;	
XX	WPI; 2001-656860/75.	
XX	N-PESDB; ABL06334.	
XX	New isolated nucleic acid detection reagent for detecting 1000 or more	
XX	genes from Drosophila and for elucidating cell signalling and cell-cell	
XX	interactions -	
XX	Disclosure; SEQ ID NO 13485; 21pp + Sequence Listing; English.	
XX	The invention relates to an isolated nucleic acid detection reagent	
XX	capable of detecting 1000 or more genes from Drosophila. The invention is	
XX	useful in developmental biology and in elucidating cell signalling and	
XX	cell-cell interactions in higher eukaryotes for the development of	
XX	insecticides, therapeutics and pharmaceutical drugs. The invention	
XX	discloses genomic DNA sequences (AB16176-AB130511), expressed DNA	
XX	sequences (AB101840-AB16175) and the encoded proteins	
XX	(AB55737-AB92072).	
XX	The sequence data for this patent did not form part of the printed	
XX	specification, but was obtained in electronic format directly from WIPO	
XX	at ftp.wipo.int/pub/published_pct_sequences.	
XX	Sequence 2441 AA;	
XX	Query Match 4.9%; Score 294; DB 22; Length 2441;	
XX	Best Local Similarity 19.8%; Pred. No. 8.8e-11;	
XX	Matches 246; Conservative 145; Mismatches 437; Indels 414; Gaps 48;	
XX	7 EISLERVAFECMGROWKHVODKLTCSHLNIEOOPAIAGS-----48	
XX	1240 EAAANSVSTNSOGKKKKQAKKKV--ISVEPTQPVTTNSVLKGVCAKKHRAVEYVKQ 1296	
XX	49 -----EDEPSQYNHSSKEISQSNPNHCKTENHR-----LEQOHNGS 84	
XX	1297 PPATQQAAPLRQLQDVKKKEEPALKKEKKSSSSSSSKREKENLAPKEVALPAKQPS 1356	
XX	85 QLEBEDSENNQTHSDSRTPTPGATSPSPRPPIRDMRPAKKNFCVN-GRLLTVNAOG 143	
XX	1357 SKLOSSEASINNS--STATNTSSANTTRKEVAPASQTASAT--TLNPAKRTVEDGWK 1411	
XX	144 KLVAEASAA-----TATSGSTS-NSHIOH--DSDNSGSASLPHHISSSS 185	
XX	1412 EVAKKSSAQOQTTAVGASGAPLPVATTSATISVQHHPHHHLANSSNSSSLTTGTTAAS 1471	
XX	186 SNNSSGNNR-----ARHTIAAASARATPAATPAANSLELYKLLTORPAKMTSDSN-- 235	

```

Db 1472 SVPEWTCCKVQVPAVNAISRIVIGRGGSNINAIKATGAHIEVEKQKQNSERCITKGLTD 1531
Qy 236 AAQALQFLLA-----DNLINSLASQOOOQOOQOIAAVNP-----TTSVSA 281
Db 1532 ATKQKMLLILKIDPDVILQMLPRINSSIKQSSGASGATPWSVGTWDRFAAGVNAVY 1591
Qy 282 ISPALKDPSPSPVDAPLDSKSPNSISGDPV-----SVRACATPTSPGRAYSEE 334
Db 1592 FSSMASTT---STSSSSASSTTPAGASYSNAHKHQOQPOSVKG-----PGRGSTS-- 1641
Qy 335 DLSRALQDVANKLDARKSASQHEQRSILDNRLEFKMKHDOEQDHG-----DELED 387
Db 1642 -----VKNSGSGSTKVSASGSGSRGRAGSYLAQOQOPGRSSGSGSSNGVIXKSES 1693
Qy 388 SNDQAEAEVDNA-----STPVYPA--EFARQRLKLSLSEING-----SD 427
Db 1694 SSKSLPAQOKSSTLIGKSSSTVSPGAONPAKAAIQQSPKAEAGATSAVVTAGGRSG 1753
Qy 428 LGEEDVDRGSPKMGHPACGNASANOAPASIPLDANVLLHTLMLAAGICAMPKLDQTV 487
Db 1754 VVAPRGKRPVAGCGGPAATAASNVAQLGVSNGENI-----LAGPIG-----TFNV.1800
Qy 488 GDFIKGLIVANSGIMNEGILNLSAQENSN-----GNASLLLQOQOHOQ 533
Db 1801 AD-VAANVAAAAAG-----AAATSNVVKPIAPAPPSKRVGSPQVQOQHOQ 1849
Qy 534 HHQCH-----HQOQOQOQOHAAYHRRLPKSTPTTNSLDPNDSEPI 577
Db 1850 QOQOQOLPOPAPVPGPOQOPLQOQOQOQO-----APQOQOQOQOQOQOTSQOQNLV 1901
Qy 578 LKIPFVKVSGPASSSSLSPGLVGGHHPLNNSNLSISNNSNHSNHSNRSRPSHA 637
Db 1902 I-----NTNLL-----1907
Qy 638 SPMILAAVAQGYSAAGNSLLTSSSSSIQKMAASNIQROIENGESLRNGVSDCSSNN 697
Db 1908 NDLMASAA-----NTTSDSFAQLAKASSAYSLFSDVQSQWQKLG-D-PCIG 1955
Qy 698 GSSSLGYKKPISIVAKIIGCTDRFGASPYLLS-----QCHSHHLLT-----HQ 744
Db 1956 GGAGVVGGLPQADASKAP-----GYNRNTLSSPVGSSKASSNHSSTSPVGNVIOQ 2007
Qy 745 QOQOQOLSAQEA-----GKGTPRKGYRNYDRDSLVEAVKAVQR-----GEM 787
Db 2008 QOQOQOSSQALNITTSGBGPAATAPAS-----PMVGBANBNPAVQGPBMNGTGLGET 2062
Qy 788 SVHRAGSYVGVPHSTLEYKVKERHLMRPKREPKQPDVLGLTGPANKLQDLKLAGPHG 847
Db 2063 APASHPGVIKPPTATVPF---QRHVPMPISAPBEAGAPPTFGAIG-----SNPAS 2108
Qy 848 GSKLSNALKNONNAAAAAATAATPAGLPLFEA-----GPO-- 891
Db 2109 G-----NNSHAAQAAAAAASAMI DROOONLQNLQTLQNLQNVGASQOQOPOQ 2158
Qy 892 -----ALSFQNMFWPTQNA-----TNAYGLDFNRTTEAMR 922
Db 2159 LNYPMWPTSSFIVDANNVLRNLPRIIPFGNKTKPPQPPGGTQGVNG-----GNGR 2212
Qy 923 NPOASNNHGLKMSAODMVENYDGI-----RKTLOASBGNGAAGNGSNGSNGNGH 975
Db 2213 QPPGT---GAROPGAAARWYGGTLEYPSYGRDMLHLENGAGAGMACGSPSAMS PMHD 2269
Qy 976 HGHGHALLDQDLVKTPLPFTNRRNDYATCASSAGESV 1017
Db 2270 D-----IRKMPRPITGERAASWKYNNFNVGPSL 2298

```

RESULT 10  
 ABB61959  
 ID ABB61959 standard; Protein; 1436 AA.  
 XX  
 AC ABB61959;

```

XX 26-MAR-2002 (first entry)
DT Drosophila melanogaster polypeptide SEQ ID NO 12669.
XX DE
XX Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX KW Drosophila melanogaster.
XX OS
XX WO200171042-A2.
XX PN
XX 27-SEP-2001.
XX PD
XX 23-MAR-2001; 2001WO-US09231.
XX PE
XX 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA
XX (PEKE ) PE CORP NY.
XX PI
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WI: 2001-656860/75.
XX DR N-PSDB; ABL06062.
XX XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX PS
XX Disclosure; SEQ ID NO 12669; 21pp + Sequence listing; English.
XX CC
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB160511), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins
XX (AB57373-AB572072).
XX CC
XX The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX Sequence 1436 AA;
XX
XX Query Match 4.9%; Score 293; DB 22; Length 1436;
XX Best Local Similarity 20.1%; Pred. No. 4.9e-11;
XX Matches 232; Conservative 132; Mismatches 350; Indels 440; Gaps 48;
Qy 206 TPAATPANSDELTKLTLTORAK-----MTSMDMAAQLAQFSLADFNLSASQOQ 259
Db 415 KTKSGSSGSLQROOALRGAVRGAMOCIFESHGHPAQLA-----NCFATPRELQ 466
Qy 260 QOQOQOIASAVTPTTSEVA-----AAISPALDTPS--PSVDAPLDLSKSPNS 308
Db 467 QLRQDMADNL-----ISASIFLPLCPALISPLSFLNITSELSARATRLTLV----- 515
Qy 309 SISGDVSVRACATPTSPGGRAYSEEDLSRALQDVANK-----DARKSASQHEQR 361
Db 516 -----AKTLQTLAFT---RFQGENFMERLNFLEQEARMOQFLEITSTRHPAPD 566
Qy 362 SILD-----NRLFQMKH-----DOEODHGDGDESDNDQAEVDNASTPVY 405
Db 567 SILDMAGYIDQGKQLSIHSLSLSLAKLPPARQHELDPLQIILDEFISRAEHGKTRL- 625
Qy 406 PAEFARQRLKLSLSEH-----NGSDLGEDVDGRSPKMGHPACGNASANOAPAS 457
Db 626 PGVILPATNSSTHSLASENGEQRNPGSSGSHGNSQLLPQSQ-----LAQGHAIIVS 679
Qy 458 IPLDANVLLHTLMLAAGICAMPKLDQTVGDPFKGLIVANSGIMNEGILNLSAQEN 517
Db 680 KPLSAB-----RGI-----MRGVLTPTS---LEK---NIFRYNDPT 709

```













**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 12:26:02 : Search time 32 Seconds  
(without alignments)  
1540.380 Million cell updates/sec

Title: US-10-016-768a-10

Perfect score: 6030

Sequence: 1 MHISSEYISLSEVSECMGR.....ILHEKLAQIKAEVDADQL 1165

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfillseq.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	252	4.2	2137	4 US-09-134-001C-4463	Sequence 4463, App
2	243	4.0	1507	4 US-09-914-259-37	Sequence 37, App1
3	231	3.8	1093	3 US-08-545-860D-55	Sequence 55, App1
4	231	3.8	1093	5 PCT-US94-04496-55	Sequence 55, App1
5	229.5	3.8	2843	1 US-07-741-940-2	Sequence 2, App1
6	229.5	3.8	2843	1 US-08-289-548A-2	Sequence 2, App1
7	229.5	3.8	2843	1 US-08-452-654-2	Sequence 2, App1
8	229.5	3.8	2843	2 US-08-370-235A-2	Sequence 2, App1
9	229.5	3.8	2843	4 US-08-449-731-2	Sequence 2, App1
10	223	3.7	1402	4 US-09-125-635-12	Sequence 12, App1
11	221.5	3.7	2842	1 US-07-741-940-7	Sequence 7, App1
12	221.5	3.7	2842	1 US-08-289-548A-7	Sequence 7, App1
13	221.5	3.7	2842	1 US-08-452-654-7	Sequence 7, App1
14	221.5	3.7	2842	4 US-08-449-731-7	Sequence 7, App1
15	221.5	3.7	2843	4 US-08-452-655B-2	Sequence 2, App1
16	221.5	3.7	2843	1 US-08-452-655B-7	Sequence 7, App1
17	221.5	3.7	2843	3 US-08-450-582-2	Sequence 2, App1
18	221.5	3.7	2843	3 US-08-450-582-7	Sequence 7, App1
19	221.5	3.7	2973	2 US-08-821-355A-7	Sequence 7, App1
20	221.5	3.7	2973	2 US-09-003-687A-7	Sequence 7, App1
21	221.5	3.7	2973	4 US-09-136-605-7	Sequence 7, App1
22	218	3.6	10182	4 US-09-134-001C-3159	Sequence 3159, App
23	214	3.5	1088	4 US-09-233-857-13	Sequence 13, App1
24	213	3.5	1099	4 US-09-442-100-2	Sequence 2, App1
25	213	3.5	1099	4 US-08-939-106-2	Sequence 2, App1
26	211.5	3.5	1752	4 US-09-556-877-180	Sequence 180, App
27	211.5	3.5	1752	4 US-09-620-412C-180	Sequence 180, App

28	211.5	3.5	1752	4 US-09-598-419-180	Sequence 180, App
29	211	3.5	2703	1 US-08-185-432-19	Sequence 19, App1
30	211	3.5	2703	4 US-08-899-232-4	Sequence 4, App1
31	204.5	3.4	1581	3 US-09-110-517-2	Sequence 2, App1
32	197	3.3	903	2 US-08-853-310-2	Sequence 2, App1
33	193.5	3.2	1420	4 US-09-125-635-4	Sequence 4, App1
34	193.5	3.2	1664	1 US-09-599-652-2	Sequence 2, App1
35	193.5	3.2	1664	2 US-08-642-846-2	Sequence 2, App1
36	193.5	3.2	1664	4 US-09-264-604-2	Sequence 2, App1
37	186.5	3.1	951	4 US-09-125-635-8	Sequence 8, App1
38	182.5	3.0	369	4 US-08-061-376-5	Sequence 5, App1
39	182	3.0	842	5 PCT-US96-02311-15	Sequence 15, App1
40	182	3.0	1326	4 US-09-147-236-5	Sequence 5, App1
41	182	3.0	1326	4 US-09-522-474-5	Sequence 5, App1
42	180.5	3.0	1464	3 US-08-891-640-2	Sequence 2, App1
43	179.5	3.0	1257	3 US-09-220-641-3	Sequence 3, App1
44	178.5	3.0	760	1 US-08-195-152-2	Sequence 2, App1
45	178.5	3.0	1776	4 US-09-556-877-179	Sequence 179, App

#### ALIGNMENTS

RESULT 1

US-09-134-001C-4463  
Sequence 4463, Application US/09134001C  
Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

CURRENT FILING DATE: 1998-08-13

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4463

LENGTH: 2137

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4463

Query Match 4.2% Score 252; DB 4; Length 2137;

Best Local Similarity 17.1% Pred. No. 2.5e-09;

Matches 201; Conservative 208; Mismatches 539; Indels 230; Gaps 33;

QY	51	EPSQYHSSKESISQ---	NPNHCKTBNRLQOHNGSOLLEB-----	EDSENNOTSHD- 100
DB	852	QASESASTSKOLSEASSTSDSASARKSESTSKTSLESTSVSDSASTSSEA	911	
QY	101	SSRTPTATSPSPPEPPIDMRPAKCNFCVNGRLVLVNAOGKLVASAAATATSSSTSN	160	
DB	912	STSTSVSSSTSTSDSTSTSDSTSDSASIKASEA-----	STSKLSASVSTSTSDAST 965	
QY	161	SHIHQDSDNS-SASLPHNIS-----	SSSSNNNSGNRAHIAASARATPAATPAN 214	
DB	966	S---TSVSDNSASTSLKSTSTSVSDSTSTSDSASTSSESDASASTSLSESTST	1022	
QY	215	SLELYKLTORAKKTTSMDSMAAQPSLADFNILNSLASQOQOQOQOQOQIASAIVPTT	274	
DB	1023	VSDSTSTSTSDSASMSASESES-----	NSKSTSLSESTSTSLSGSTAST 1067	
QY	275	SEVSAATSPALKPDPSPVDAPIDLSSKPPNSISGDVSVACATPPSGRAYSEE	334	
DB	1068	SDSASTSESESDSTSL-----	SESTSTSLSGST-----ASTSDASTSTSES 1114	
QY	335	D-----LSPALQDVANKLDARKSASQHEQRSIIDNRLFKKXHHDOEDHGDDELED	387	
DB	1115	DSTSESTSLSESLSTSVSDSTSDASTSESA-----	STSTSESENSASTSLSG 1161	

```

OY 388 S-----NDAAEAEVDSNASTPYRPEAPFAQURKLSHSEHNGBSLGVDURGSPKMRH 442
Db 1162 SLSTISIDSTSTSDTSDASTSSEESDOST---TSISESTJSLSDSTSTSESAST 1217
OY 443 PACGNASANOGA PASIPIPDANVLHTLTLMAAGIGAMPKLDETOTUGDETFKLLVANSAGI 502
Db 1218 STSEEDSTSEESTJSLSESTSVSDST-----SASTSDASSTSVSD-----SESASTS 1266
OY 503 MNEGLNLILSASQENSNGNASILLLOOOHOCHHOQHOOHQOOOQOHVAAYRRHLPKSETPE 562
Db 1267 ISESLSTSVSDSTSTSDSA-----STSTSE 1293
OY 563 TNSSLDPADASEDPILIKIPFKVSGPAPASSLSFG-----GLVGHHPLNNNSL 613
Db 1284 SDSTSESTSLSESTSVSDSTSDASTSDASTSESTSESDASTSLSGSTJSLSDST 1355
OY 614 SISNNSNH-----SNSHRNGSNRPSPHASPMIAAVALAQGYASGNLSLTSSSGSIOKMAS 670
Db 1354 STSDSASTSESDDERASTSLSGSTJSLSDSTSTSDASTSVSDSASTSVSDSASTSLSG 1413
OY 671 NIQRINOINQSODES--LANGNVSDCSSNNGGSSSLGYKKPSIVAKITGGDTSPFQASP 728
Db 1414 SLSTSVSDSTSTSTSDASASTSESDSRASSTLSGSTJSLSDST---STJSDSASTS 1470
OY 729 NILSGOHSAAHMLTHQQOOOOLASQEALEGKTGRPRKGXRYN--DRDSL----VEAVKVQ 783
Db 1471 TSVSERNST---TSISESLSTSVSDSTSTSDASTSVSDSDASTSSSESVSD 1527
OY 784 RGENSVFRAGSYGVPHSTLEKYKERHLMPKRKEPRORDVLGT-----830
Db 1528 SESTSTSTSDS-----ASTSTSVESN-----STJSLSGSTSVSDSTSTSTST 1572
OY 831 --GPANKUKQLKLKAPRGGSKLSALKNONNOAAAAMAAAAATPNCGLKPLFEA 888
Db 1573 DSABASTSESDSDASTSSSESVSTSV--SDTSMSTSESASTSVSDSNASTSLSES 1630
OY 889 GPQALSFPQMFWPQTNAUYGLDFNRITTEAMRPQASNHHGLMKSAQDWENYVDGI 948
Db 1631 TSTSLSDSTMSTSDASTSESDSDASTSLSDSTST-----1670
OY 949 KRTLOASGNGSAAANGSNGNGHGCHGHGHALD--OLLVKTRLPPTNHRNDYA 1006
Db 1671 -----VSESTSTSTSVASANNSTJSLSDRSTJSLSDSTSTSESGSTSESDSPA 1725
OY 1007 AT-----CSASAGEVKSRCSPMNAYDIKERS-----ADSGSGSESHS 1048
Db 1726 STJSLSESTJSLSDSTSTSDASTSKMSVSDSNDAJSLSDSTSVSDTSAJSTESA 1785
OY 1049 ASHINNNSDLAHNKXKSGGGGGCGGCGGQOTONGNRSRMTGRDDSETDPASSFSGENGQ 1108
Db 1786 STSTRESSTASTSLSESTSVSDSTSTSTSDASTSESDNSESTJLSESTSV 1845
OY 1109 QNHQMMDLNGSSSSSHIKCESEAATGHSNGHNTSI 1146
Db 1846 SDSTSAJSTASTSVSDSNASTS--LSGSTSV 1881

```

```

RESULT 2
US-09-914-259-37
; Sequence 37, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 1507

```

```

; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-914-259-37

```

Query Match	4.0%;	Score 243;	DB 4;	Length 1507;
Best Local Similarity	18.6%;	Pred. NO. 6.9e-09;		
Matches 241;	Conservative 157;	Mismatches 433;	Indels 462;	Gaps 48;

[illegible]



```

Db      1197 MAKVNLV-PQOOQHGNKRHLNLSATGAGNPVBSKRLSGCTLGDVQSPQLLQGLIGKDPAQ 1255S
Qy      858 OMNQAAAAAAAAAAAAAT-----PGLKLPFEAG-----PQALSFQPMFWPQ 903
Db      1256 QOTPAARKAGERWOLSAESKQOKQOQOQSSVYLKNLLVSGRDDDDSEAMIIIDENSLVQ 1315S
Qy      904 TNATNAYGLDNRRLITEARNRPOASHHGLMKSADQMVENVYDGIIRKTLQASGNGSAAQ 963
Db      1216 PIPGLKGYLPLCHCTSTSVLRDYNHNPILIGTFNQLSPVIG-----GSDSGG 1364S
Qy      964 NGSNGSNGNGHGHHGHGHALLDOLLYKTPPLFTNHRNNDVYAATCSSASGESVYKSGSP 1023S
Db      1365 DGETGGSVVS-----LDSDSVPPGLT-----ACDTPDASSDSGIDENSL 1400S
Qy      1024 MGNVADIRERLSDSGSSDPEHSASHIINNNSDILAHNKAKSGGGGGGGNGQOTNGNR 1083S
Db      1401 MDGASGSFRKRLSSTNSNTQNAESAIPPALDVETPVTKQSVEEFEGGSGSNA----- 1453S
Qy      1084 SSRMTSRDSETFDASSFKSGENGGAQNHKMMDL 1116
Db      1454 PSRKTS-----ISFLDSSNPILLHTPAMMDL 1478

```

RESULT 3  
 US-08-545-860D-55  
 Sequence 55, Application US/08545860D  
 Patent No. 6040140  
 GENERAL INFORMATION:  
 APPLICANT: Croce, Carlo  
 APPLICANT: Canaan, Eli  
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1 Region  
 NUMBER OF SEQUENCES: 94  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &  
 STREET: One Liberty Place, 46th floor  
 CITY: Philadelphia  
 STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19103  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/545,860D  
 FILING DATE: 07-MAR-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US94/04496  
 FILING DATE: 22-APR-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US92/10930  
 FILING DATE: 09-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/327,392  
 FILING DATE: 19-OCT-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/320,559  
 FILING DATE: 11-OCT-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/062,443  
 FILING DATE: 14-MAY-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/971,094  
 FILING DATE: 30-OCT-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/888,839  
 FILING DATE: 27-MAY-1992

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07//805,093  
 FILING DATE: 11-DEC-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Deluca Esq., Mark  
 REGISTRATION NUMBER: 33,229  
 REFERENCE/DOCKET NUMBER: TLU-1262  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 568-3100  
 TELEFAX: (215) 568-3439  
 INFORMATION FOR SEQ ID NO: 55:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1093 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-545-860D-55

Query Match	3.8%;	Score 231;	DB 3;	Length 1093;
Best Local Similarity	19.9%;	Pred. NO. 3.3e-08;		
Matches 206;	Conservative 116;	Mismatches 373;	Indels 342;	Gaps 40;

[illegible]

```

OY 634 PHSAPMLAAVAAGGYASAGNSLITSSSSSIQKMA-----NIQPIINGSGEISR 686
Db 830 FHSTPEPL-----PLIQSPPTITLALPGAPAPLPPOPQNGLGAPAGAIG- 875
OY 687 NGNVSDCSNNGSSSLGYKKPSISVAKIIG-----TDTSRFGASPVL-----LS 732
Db 876 -GAMMAEGLGLAGSG-----GLPLNGILGLNGMAAHPNPSLSQAGAPITLQIPGLIN 930
OY 733 QOHSAHHLTHQOQOQOOLSAQEBALCKGTRPKRGKYRNYDRDSLVEAYAVAGRGEMSVHRA 792
Db 921 SLTEGQRHLLQOQEQOQLQOQL-----A 955
OY 793 GSYIVVPHSTLEYKY----KERHLMRPKPEKPKQPDIVGLTGPAINKLQDKLKAGPHG 847
Db 956 SPQLTPEHQTVVQMIQIQKRELFQRIQMAQSGLPMASLLAGSSTPL-----LSAGTPIG 1011
OY 848 -----GSKLSNALKNQNNQAAAAAATAATTPNGCLKPLFEAPQALS 894
Db 1012 LLPTASAPPLLPAGALVAPSLGNNTSLMAAAAAAAAVAA-----GGPPIVLT 1056
OY 895 FQPNMEWPQTNATNAYG 911
Db 1059 AQTNPFLLSGAEGSGG 1075

```

## RESULT 4

```

PCT-US94-04496-55
: Sequence 55, Application PC/TUS9404496
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
: TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
: NUMBER OF SEQUENCES: 86
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
: ADDRESSEE: Norris
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04496
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca Esq., Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: TUD-1242
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ. ID NO. 55:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1093 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
PCT-US94-04496-55

```

Query Match	3.8%;	Score 231;	DB 5;	Length 1093;
Best Local Similarity	19.9%;	Pred. No. 3.3e-08;		
Matches 206;	Conservative 116;	Mismatches 373;	Indels 342;	Gaps 40

CY		19	GRQMKHYODLCTGSHLNIEQQCI-----AIAGEBDRPQYNHSSKXISQGN-	66
Dd		237	GOKSRDKDEHLKKQHKRRPPSPSILTPPVVPADKYSSASSSSHENASTOTSESR	296
OY		67	-PNHKTENHLEOQHNGSOLEBED-----SENNOTSHDSRPTPGATST-----	112
Dd		297	ESKGKSSSHSL-SHKQKLSSGKVGSFTASSSSSSSSGCFQPQAVSLQSSPD	354
OY		113	-----BPPEPEPIDMRSAKNFCVNRRLLTVAAGCLVAESA	150
Dd		355	FSAPFKLEQPEEDKYSKTAPAPAPPSPSAPEPKDLF-----EQVVVPSGF	403
OY		151	ATATSSSTNSNHIHQHDSDNSSALSPLHHISSSSNNNSGNRAHIIAASARATPAAA	210
Dd		404	GPIMRFSTTS-----SSGARARAPSDGYKSPIVYTSGASAGTHKMXPALSTATPVPADE	457
OY		211	TPANSLELYKLLTPRAAKMT-----SNDSMAAOIAOTSLLADF	248
Dd		458	TPETGLKEKK--HKASKRSRHGGPRKSGRNKEGTGPAPAPSLPS--AQJAGETAATA--	510
OY		249	NLINSLSAQOOQQOQQOIASAVTPTTSEVSAAAI SPLKDTPSPVAPL-----DLSKP	304
Dd		511	-----ASPBSG-----GSLVSSGLGOLSRTTGPS-GSLPSLSLSPLLGAGITYSNK	557
OY		305	SPNSISGDYKSVACACTPT-----PSGRAVEEDLSRALQDVANKLDARKSASQHN	358
Dd		558	DPISHSGMLAAV--CSTPLASSSLGPQTSA-----LPRLRSRPFSTLPSSEA----	605
OY		359	EQRSLINDRLFKMHNHQEOQHDDLEBDSNDDAEAVDSNASTPPVPAEFARQRLKS	418
Dd		606	--SISTTVVFSL-----AGSTFSLPHTHIFGTPMGAVNPILLSQ	641
OY		419	HLSHENGSDELCEVDRCSPKWRHPACNASANOGAPASIPLDANVLHTLMLAIGAM	478
Dd		642	AESHTEPDEDLDCSFR-----CRGTSPOQSLSMSP-----ISSL	676
OY		479	PKL-DETQTVGDFTIKGLLVANGGIN---EGLINLSAOGENSNGNASILLQOOQHQH	534
Dd		677	PALPDQTA-----APCGGOQLDPAAPGTTNMBOULEKOQDGEGAVNIEMLKALH	727
OY		535	HQOHHOQOQOQ--HVAAYRHRLP-----KSETPETNS-----SLDPNDAS	573
Dd		728	ALQKENQRLQEQLISLTKKERLOQLINVOLSVFPALPALPANGVPRGYGLPQAGS	787
OY		574	EDPLIKTFSPKVSPPASSSSLSPGVLGVGHHPPLNNNSLSISNNSHSSHRNGSNRS	633
Dd		788	SDSL-----STKSPPG-----KSLIGLDNSLSTSEDPHSCGPSRSSSLS	829
OY		634	PHSASPMLAAVAOCGYSAAGNSLLTSSSSI QKMMAS-----NIQOIINEQGQEBLR	686
Dd		830	FHSITPPPL-----PLLQOSPATLPLALPGAAPALPPOPONGICLRAGAAGL-	875
OY		687	NGANYSDCSNNGSSSLGYKKPSISVAKTIIG-----TTSRFQASENL-----LS	732
Dd		876	-GAMPABGILLGLAGSG----GIPLINGLIGLNGAARNPASLSQAAGATLLDPCGLN	930
OY		733	QOHSASHLHTQOQOQOQLSAOELGCKGTRPKRGKRYNYDRDSLVAEYKAVOGENMSHRA	792
Dd		931	SLTEOORHLLOODEOLOLOOLL-----A	955
OY		793	GSYYGVPHSTLEYUV-----KERILMPRKREKPROPDVGLTGPANKLQDLKAKBGHG	847
Dd		956	SPQLTPEHQIVYQMIOIQOKRELQLOMQNAGSQULPMASILLAGSSTPL-----ISAQTPG	1011
OY		848	-----GSKLSNALKNONNOAAAAAAAAAAAAAATENGKLPLEAGPOLAS	894
Dd		1012	LLPTRASAPLLPAGALVAPSLGNTTSLMAAAAAAAAAAAAAA-----GGPVLIT	1058
OY		895	FQPMFMFOQTNATNAYG	911
Dd		1059	AQTNPFUSLSGAEESGG	1075

RESULT 5  
US-07-741-940-2  
Sequence 2, Application US/0741940  
Patent No. 5352775  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THILVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/741,940  
FILING DATE: 19920109  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.035574  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-741-940-2

Query Match 3.8%; Score 229.5; DB 1; Length 2843;  
Best Local Similarity 19.3%; Pred. No. 1.6e-07;  
Matches 266; Conservative 179; Mismatches 510; Indels 421; Gaps 63;

2 HISSEIISLERYAECEMGRRO---WGYODKLTCSHLNIE---EQOPIA-----IAGS 48  
1054 HIIEBEI---KQSEKQOSNOSTTYVYTESTDDKHLKQPHGQOECSPFRSGANGS 1110  
49 EDEPQVYHS-SKEISQSNPHNCKTENHRLLEQHNQSQILLEEDSENNQTSHTSRPTP 107  
1111 ENRVOSNNGINQNSQS---LQEDDYEDDKPTVYSERYSEEE---QHEEBERPTN 1161  
108 GA---TSTPSPPEPIDMWRPSAKNCNVCVNGRLITVNAQKLVAESATPTSSSTNSHH 164  
1162 VIKTNEERKRVHDQPTDY-----SLKVTDTIPSSQKOSF 1195  
165 QHDSNNSASLPHHISSSSSSSNNNSGNRAR---HIAASARA---TPAATPANSLE 217  
1196 SFSKSSSSGSSSTEHSSSSSESTSTSSNAKQONQHPSSAOSRSGQPPKAAITCKVSSIN 1255  
218 LYKLLTORAAKMTSMDMAQLAQSFLADFNILNSLASQOQOQOQOQOIA-----S 268  
1256 QETIQYCVEDTPICFPSRCSSTLSLSAEDRIGCNQTTQEADSAANTLQIAETKIGKIGTRS 1315

269 AVPTTSEVSAAPALPKDTPSPSVDAPLDLSKXPSNNSISGDKVSVACATPTPSGR 328  
1316 AEDP-VSEVPVAVSQHPRTK-----SRLOGSSISSE--SARHAAVFPFGA 1358  
329 RAYSEEDLSRALQDVVANKLDARKSASQHEORS-ILDNRLEFKMGHD-----Q 376  
1359 KSPSK-----SGAQPKSPPEHYQETPLMRSCSVSLSPFSRSIASSVQ 1406  
377 BODHDG-----DELEDSDDAEVDSNASTVYVAEFAPRAQLRKLSHLSEHNSDL 428  
1407 SEPCSGMVSGIISPSDLPS--PGQTMPPSSSKTPPPPTQATOK-REV---PKKKAPT 1459  
429 GEDVDRGSPKGRPHACGNASANOGAPASIPIDANVLH--TLMAAGIGAMPKLDERTQ 486  
1460 AEKRESG-PKO---AAVNAVOR---VQLPDAUTLHFAFTESTPPDGFSGSSLSAISL 1511  
487 VGDPIKGLLVANSGGINNEGILNLSASQENSNGNASLLQOQOQHQQHHQHQOQOQO 546  
1512 DEPIQ-----KDVLRIMPVQENDNGN-----ETESQPKESNNOEKAEK 1555  
547 HVAAYRHRLPKSE-----TPETNSGLDPNDA---SEDPIL 578  
1556 TIDSEKDLDDSDDDIELECIISAMPTRKSRGKKPAQTASKLPPVARKPQLPVY 1615  
579 K-IPSFKVSGPASSSLSPGC-----LVGGHHPLNNSLSISN-----NSNHSN 624  
1616 KLIPSQNRLQPOKAVSFTPGDDMPRVYVEG--TPINSTATSLDILTETESPDLAIGE 1673  
625 SHRNGSNRSPHASFML-----AAVAOGVSA-----GNSLTSSSSSIO 665  
1674 GVRGAQSGEPFKRPTITBEGSSTDEAQGKTSSTVITELDONKAEEDDILAEICNSAMP 1733  
666 K-----MAAGNIQIINEGQGESLRNGVNSDCSNNNGSSSLGYPKPSISVAKIIGCT 719  
1734 KQKSHKPPRVKIMQVOOASASSAPKMNQLD-----GKKKKPTSPVPXPIPN 1782  
720 DTSRGASPNLLSQOHHSHHLTHQOQOQOQASQALCKGTRPKRGKY-RVYDRSLVEA 778  
1783 TEYRTVRKADSKNLNAERFSPDNKD--SKKNLKNNSKDPNDKLPNNEDR----- 1833  
779 VKAVORGEM---SVRAGSYGVPH-----STLEKVKERHLMRKRKEPKQPDLV 827  
1834 ---VRGFAFDPSPHHYPIEGTPTCFSRNDSLSLDDDDDDVLSR----- 1876  
828 GLTGPAKQLQDKLKAGPHGSKLSNALKNQNNQAAAAAATAATPGLPLPFE 887  
1877 ---EKAELRKAKENKESAKVTSHTLTSNQSANKTQAI-----KQPINR 1920  
888 AGPOLSTFQPMKFWQTN-----ATNAYGLDPNRITTEAMRN-POASNHGLMKSAOD 938  
1921 GQPKPT-LQKOSTFPOSSKQIDPRGAATDEKQNR---AIENTPVCFSHNSLSLSLD 1974  
939 WVE---NVYDGIKRTLQASEGNGSAAGNGSNGNGHGHGHALLDOLLVKKTPPL 995  
1975 IDQENNNENEPKKT-EPPOSQEP-----SKPQASGVAKPSFH-----VEDTPV 2019  
996 PPTNHRN-----NDYAATCSAS-----GESVKSGSPWG-----N 1026  
2020 CFSRNSLSLSLIDSEDDLQECISSAMPKXKPSRLKGNBKSFRNNGGILGEDLTLD 2079  
1027 YADIKR---BRLSADS-----GSSGDEH----- 1047  
2080 LKDIDRPSERHGLSPDSENFPMKAIQEGANSIVSSLHQAAAAACLSROASDSDSLSLK 2139  
1048 ---SASHINNNSDLAHHNKNKSGGGGCGNGQOTGNGRSSRMTSRDSETPASFX 1101  
2140 SGISLGSFPHLTPDOEKPTSNK-----GRILKRGKSTLETETKISSES----- 2185  
1102 SGENGQGNHKKMDLNGSSSSSHIKCE-----SEATGHHSPGHHTTS 1145  
2186 KGIGKGVKYSKL-ITGKVRNSSEISQMKOPLAANMPSISRGRTMIHPGVNRS 2240

RESULT 6  
US-08-289-548A-2  
Sequence 2, Application US/08289548A  
Patent No. 5648212  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOF.  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THILVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti, LTD  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/289,548A  
FILING DATE: 12-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.46943  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-289-548A-2

Query Match 3.8%; Score 229.5; DB 1; Length 2843;  
Best Local Similarity 19.3%; Pred. No. 1.6e-07;  
Matches 266; Conservative 179; Mismatches 510; Indels 421; Gaps 63;

QY 2 HISSEIISLVAECMGRQ--WKHYQDKLTCSHLNIE---EQQPIA-----IAGS 48  
DB 1054 HIIIEI--KQSEOROSNQSTTYPVYTESFDKHLKQPHFGQECVSPYRSGANGS 1110  
QY 49 EDEPQVYHS-SKETISQSNPHNCKTENRLEQOHNGSOLLEBEDENNOTSISRTPTP 107  
DB 1111 ETNRVSGNHGINQNVQS--LCQEDDYEDDKPTVYSERYSEEE-----QHEEBERPTN 1161  
QY 108 GA---TSTPSPPEPIDMPPSAKCNFCVNGRLTYNAQOKLVAESAATATSSSTNSIH 164  
DB 1162 YSIKNEERKHYDQIDY-----SLKYATIDIPSSQKQSF 1195  
QY 165 QHDSQNSASLPHHISSSSSNNNSGGRAR---HIAASARA---TPAATPANSLE 217  
DB 1196 SFSKSSSGGSKTEHMSSESTSTPSSAKRQNLPHSSAQSRGQPGKATCKVSSIN 1255  
QY 218 LYKLTQRAAKTSHDSMAAQLAQFSLADENLINSLSAQOQOQOQOQOIA-----S 268  
DB 1256 QETIQTYCEDTPICFSRSSSLSSLSAEDIGCQNTQOEDASANTLTQIAETKGIQTRIS 1315

QY 269 AVPTTSEVSAASIPALKDTPSPSDVADPLDSSKSPSSISGQDVXSVRACATPTPSGR 328  
DB 1316 AEDP-VSEVPVAVSQHPRK-----SRLQSSLSSE--SARIKAVFEPGCA 1358  
QY 329 RAYSEEDLSRALQDVANKLDARKASQOHEORS-IIDNRLFKMKHND-----Q 376  
DB 1359 KSPSK-----SGAQTPKSPPEHYQETPLMFRCSTSVSLDSFESRSIASSVQ 1406  
QY 377 EODHDG-----DELEDSNDDAEAVDSNASTPVVPAEFARQRLRSLSHSNGSDL 428  
DB 1407 SEPQSGWVSGIISPSDLPDS--PGQTMPSPSKTTPPPPOTAQTK-REV---PDKAAPT 1459  
QY 429 GEDVDRGSPKMGHPACGNASAGAPASIPLDANVLH--TLMAAGIGAMPKLDQETOT 486  
DB 1460 AEKRESG-PKQ---AAVNAVQR---VQVLPADTLHFTFESTPPOGSCSSLSALSL 1511  
QY 487 VGDFTKGLLVNAGGINNEGILNLSAQENSNGNASILLQOQOQHQQHQQOQOQOQ 546  
DB 1512 DEPFIO-----KVELRIIMPVOENDNGN---ETSEOPKESNENQEKAEK 1555  
QY 547 HVAAYRHRLPKSE-----TPEITNSLDPNDA---SEDPIL 578  
DB 1556 TIDSEKILLDSDDDDIIELECTISAMPTKSKKKKPAQTAASKLPPVARKESQLPVY 1615  
QY 579 K-IPSFKVSGPASSSSLSPGQ-----LVGHHHPLNNNSLSISN-----NSHSSN 624  
DB 1616 KLTPSQNRLOPQKQIVSFTPGDDMPRYCVES--TPINPSTATSLSDLTIESPRELAGE 1673  
QY 625 SHRNGSNRSPHASPMI-----AAVAQGYSA-----GNSLITSSSSSIQ 665  
DB 1674 GVRGAGQSGEPEKEDTIPTEGRSTDEAGQKTSVTLPELDNKAEEBDILAEICINSMP 1733  
QY 666 K-----MMAINIQROINEQSGESLNGVNSDCSSNNGSSSIQYKPPSI-SVAKIIGCT 719  
DB 1734 KKKKHPKPVKKINDQVQOASASSAPVKQLD-----GKKKKPTSPVKPIPON 1782  
QY 720 DTSFEGASPNLISQOHSANHLTHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 778  
DB 1783 TEYTRARKADNSGNLNAERVFSDND--SKQNLKNNSKDNDKLPNNEDR----- 1833  
QY 779 VKAVQGEH---SVHRAQSYGVPH-----STLEYKVERHLMPRKREPKPODLY 827  
DB 1834 ---VRGSEFAPDSPHHYPIEGTPYCFSRNDSLSTLFDDDDDVLSR----- 1876  
QY 828 GLTGPANKLODKLKAGHGGSKLSNALKNQNNQAAAAAATAATPNGLKLPFE 887  
DB 1877 -----EKAELRKAKENKESPAKVTSHTELTSNQOANKTOAIA-----KQPINR 1920  
QY 888 AGPOLSEFQPMFWPQTN-----ATNAYGLDFNRITTEAMRN-POASNHGLMKSAD 938  
DB 1921 GQPKPI-LQKXSTPQSSKQIDPRGATDEKQNF-----AIENTPVCFSHNSLSLSLD 1974  
QY 939 WVE---NVYDGIKKTLQASEGNSAAGNSGNSGNGHGHGHAALLDQLLYKKTPL 995  
DB 1975 IDQENNNKENPIKET-EPPDSQGEF-----SKPOASGVAAPKSFH-----VEDTPV 2019  
QY 996 PETNHRN-----NDYATCSSAS-----GESVKKSGSPMG-----N 1026  
DB 2020 CFSRNSLSLSISIDSEDDLOECTISSAMPKKKPSRLKGDVKKSPNNMGILCEDLTLD 2079  
QY 1027 YADIKR-----ERLSADS-----GGSDEEH----- 1047  
DB 2080 LKDIQRPSEHGLSPDSENFPMKALIQEGANSIVSLHQAALAAALSRQASDSISLSLK 2139  
QY 1048 -----SASHINNNSDLAHNKNKSGGGGGGNGQTNNGRSGSRMTSRDSETDASSFK 1101  
DB 2140 SGISLSPFHLTPQEEKPFTSNK-----GRIILPKGKSLTLETKLIESES----- 2185  
QY 1102 SGENGGQNNHGMMLNGSSSSSHIKCE-----SEAATGHSPGHHTTS 1145  
DB 2186 KGIGGKKVYKSL-ITGKVRNSSEISGOMKQPIQANMPISISRGRTWHIPVRNNS 2240

```

RESULT 7
US-08-452-654-2
; Sequence 2, Application US/08452654
; Patent No. 5681454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODNE, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THILVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-452-654-2

Query Match      3.8%; Score 229.5; DB 1; Length 2843;
Best Local Similarity 19.3%; Pred. No. 1.6e-07;
Matches 266; Conservative 179; Mismatches 510; Indels 421; Gaps 63;

QY      2 HISSEISLIERVAEECMGRQ--WKHYODKLTCHLNIE---EQQPIA-----IAGS 48
DB      1054 HIIIEI---KQSEOROSHNQSTTYPVYESTDDKHLKQPHFGQOECVSPYRSRGANGS 1110
QY      49 EEPEQVYNS-SKETISQSNPNHCKTENHLEQOHNGSOLLEEDSENNOTSHDSRTPTP 107
DB      1111 ETNRVGSNNGINQNSQS--LCQEDVDYDDKPTVYSERYSEB-----QHEEERPTN 1161
QY      108 GA---TSTPSPPEPIDMRPSAKCNFCVNGRLITVAQGLVAESAATATSSSTNSHIH 164
DB      1162 YSIKYNEXRHYVDQPIDY-----SLKYATIDTIPSSQKQSF 1195
QY      165 QHDSQNSASLPHHISSSSSNNNSGGRAR---HIAASARA---TPAATPAMSLE 217
DB      1196 SFSKSSGSGSSTKHEHSSSSSENTSTPSSNAKRONOLHPSSAQSRSGQOPKAAATCKVSSIN 1255

```

```

QY      218 LYLKLTQRAAKWTSMDSMAAQLAFSLADPFLNLINSLASQOQOQOQOQOIA-----S 268
DB      1256 QETIOTYCEVDPTPICFSSCSLSLSAEDLIGCQTTOQPADSANTLLOIAELKGIKIGTRS 1315
QY      269 AVPTTSEVSAASAIISPAKDTPPSPVDAPLDLSKSPNSISGCVKSVRCATPTPSGR 328
DB      1316 AEDP-VSEVPVAVSQPRTK-----SSRLGSSLSSE--SARHKAVEFPFGA 1358
QY      329 RAYSEEDLSRALQDVAVNAKLDAKASQHHQRS-ILDNRLEFKMHND-----Q 376
DB      1359 KSPSK-----SGAOTPKSPPEHYVQETPLMFSRCTSVSLSPFSRSIASSVQ 1406
QY      377 EODHDG-----DELESDNDAAEVDNASSTFVYPAEFARAOLRLKLSHLENGSDL 428
DB      1407 SEPCGWSGIIISPDLPDS--PGOTMPPSSKTPPPPPQTAOTK-REV---PKNAPT 1459
QY      429 GEDVDKSPKXGRHAPACGNASANOAPASIPLDANVLLH-TLMLAIGAMPKLDTOT 486
DB      1460 AEKRESG-PKQ---AAVNAAVQR---VQVLPDADTLHFATSTPPDGFSCSSLSALSL 1511
QY      487 VQDFIKGLLVANSGGIMNEGILNLISAQENSNGASLLLOQOQHQQHQQHQQOQOQ 546
DB      1512 DEPFIQ-----XVELRLIMPVQENDGN---ETESQPKESNENDEKAEK 1555
QY      547 HVAAYRRLPKSE-----TPETNSLDPNDA---SEDPIL 578
DB      1556 TIDSEKDLIDSDDDDIILEECIISAMPTSSRKPKPAOTASLPLPPVARKPSQLPBY 1615
QY      579 K-IPSFVSVSPASSSSSPGQ-----LVGHHHPNNSNLSISN-----NSNHSN 624
DB      1616 KLPFQNRLOQKIVKSFPGDDMPRVYVEG--PTINSTATSLDLTIESPNNLAAGE 1673
QY      625 SHRQNSNPSHSASML-----AAVAQGYSA-----GNSLITSSSSSIQ 665
DB      1674 GVRGAQSGEPEKRDITPTBGRSTDEAQGXTSTTIPELDNKAEBEDDIIAECINSAMP 1733
QY      666 K-----WMAINIOQIENSGQESLNRGNVSDCSNNGSSSGYKKPSISVAKIIGT 719
DB      1734 KQKSHKPRVKKINDOVQASASSAPYKNQD-----GKKKKPSPVKPIQN 1782
QY      720 DTSRFGASPNLLSQOHHSAHHLTHQOQOQOQLSAGEALCKGTRPKRKY-RNYDRDSLVEA 778
DB      1783 TEYRTRVAKNADSKNNLAERAVFSDNKD--SKQNLKNNSKDFNDKLPNNEDR----- 1833
QY      779 VKAVDQGM---SVHRAGSYGVPH-----STLEKVKERHLMRKRKEPKQPDLY 827
DB      1834 ---VRGSFAPDPSPHHYPIEGTPTCYFSRNDLSLDPDDDDVDLSR----- 1876
QY      828 GLTGPNKLOLDKLCAGPHGSKLSNALKNQNNQAAAAAATAATPNGKLPFE 887
DB      1877 -----EKAELRKAKENKESAKVTSITELTNSQOANKTQALIA-----KQPINR 1920
QY      888 AGPQALSFPQPMWFMPQTN-----ATNAYGLDFNRIITEMARN-PQASNHGMLKMSAD 938
DB      1921 GQPKPI-LQKSTPQSSKQDLPDGAATDEKLQNR-----AIENTPVCFSHNSLSLSLD 1974
QY      939 MVE---NVYDGIIRKTLQASGNGSAAAGNSGNGNGHGHGHGHALLDQLVYKTPPL 995
DB      1975 IDQENNKNKEPIKET-EPPSYGEP-----SKPQASGYAPKSFH-----VEDTPV 2019
QY      996 PETHNRN-----NDYATACSSAS-----GESVKSQSGPMG-----N 1026
DB      2020 CFSRNSLSLSIDSEDDLQECTISSAMPKKKPSRLKGDNEKSPRMNGGILGEDTLTD 2079
QY      1027 YADIKR-----ERLSADS-----GSSSDEEH----- 1047
DB      2080 LKDIQRPDSHGLSPDSENFPMWKAIQEGANSIVSLHQAAAAACLSRQASDSDS1LSLK 2139
QY      1048 -----SASHINNNSDLAHNKVKSGGGGGCGNGQTNNGRSSRMTSRDSEFDASSFK 1101
DB      2140 SGISLSPFHLTPQDEEKPFTSNK-----GPRILKPGKSTLETAKIESE----- 2185
QY      1102 SGENGQOQNHMMQDINGSSSSSHIKCE-----SEATGTHHSFGHHTTS 1145

```

Db 2186 KGKGGKVVYKSL-ITGKVRNSSEISGQMKPQLQANMPSISRGRTMIHIFGVRRNS 2240

## RESULT 8

US-08-370-235A-2

Sequence 2, Application US/08370235A

Patent No. 5910418

GENERAL INFORMATION:

APPLICANT: VOGELSTEIN, BERT

APPLICANT: KINZLER, KENNETH W.

APPLICANT: HILL, DAVID E.

APPLICANT: JOHNSON, KAREN A.

TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DETERMINING

TITLE OF INVENTION: MUTATIONS IN THE APC GENE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: BANNER &amp; WITCOFF, LTD.

STREET: 1001 G STREET, N.W.

CITY: WASHINGTON

STATE: DC

COUNTRY: US

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/370,235A

FILING DATE: 01-JAN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: KAGAN, SARAH A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 01107, 48688

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202 508 9100

TELEPHONE: 202 508 9299

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2843 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-370-235A-2

Query Match 3.8%; Score 229.5; DB 2; Length 2843;

Best Local Similarity 19.3%; Pred. No. 1.6e-07;

Matches 266; Conservative 179; Mismatches 510; Indels 421; Gaps 63;

QY 2 HISSYEISLERYAEBCMGRRQ---WKHYODKLTGSHLINE-----EQQPIA-----IAGS 48  
Db 1054 HIIIEI---KQSEQRQSHNOSTTYPVYTESTDDHLKFPHPGOECVSPYRSGANGS 1110  
QY 49 EDEPSQVYNS-SKEISQSNPNHCKTENHLEQOHNGSQLLEEDEENNQTSRSTPTP 107  
Db 1111 ETRVGSNHNGINQNSQS---LQEDDYEDDKPTVYSERYSEEE-----QHEBERPTN 1161  
QY 108 GA---TSTPSPPEPIIDMRPSAKCNFCVNGRLITVNAQGLVAESAATRTSTSTNSHIH 164  
Db 1162 YSIKYNKEKRVHDQPIDY-----SLKATYTDIPSSQKOSF 1195  
QY 165 QHDSQNSASLPHHISSSSSNNNSGNRRAR---HTAAGARA---TPAATPANSLE 217  
Db 1196 SFSKSSSGOSSKTEHMSSESTSTPSSNAKQNLHPSSAOSRSGOPKAAATCVYSSTIN 1255  
QY 218 LYKLTORAAKNTSNDMAAQLAFSLADFNLIINSLSAQOQOQOQOQOIA-----S 268  
Db 1256 QETICTYCEDTPIPCFRSSSISSLSADELICGNOTTOEASANTLOJAEIKKIGTRS 1315  
QY 269 AVTPTTSEVSAAIIPALKDTPSPVDADPLDISSKPSNNSISGVKSVRACATPTPSGR 328

Db 1316 AEDP-VSEVPVAVSQHPRTK-----SSRLQGSLSSE--SARRKAVEPBGCA 1358  
QY 329 RAYSEEDLSRALQDVANKLDARKSASQHEQRS-IIDNRLFPKKKHHN-----Q 376  
Db 1359 KSPSK-----SGAOTPKSPPEHYQETPLMFSRCTSVSSLSDFSRSLASSVQ 1406  
QY 377 EQDHDG-----DELEDSDDAEAEVDSNASTPVVYPAEPARAQLRKLSHLSEHNSDL 428  
Db 1407 SEPCSGWVSGLISPSDLPDS--PGQTMPSPSKTTPPPPPQAQTK-REV-----PKNKAPT 1459  
QY 429 GEDVDROGPKXGRBPACGNASANOAPASITLDANVLH--TMLAAGIGAMPULDETQOT 486  
Db 1460 AEKESG-PKQ---AAVNAVQR---VOYLPDADTLHFATSTPDGFCSSSLALS 1511  
QY 487 VGDFIKGLLVANSGIMNEGILNLSASOENSNGNASILQOQOQHGHQHOQOQOQO 546  
Db 1512 DEPTIQ-----KDVFLRIMPVQENDNGN-----ETESQPRESENENQEKAK 1555  
QY 547 HVAAYRRRLPKSE-----TPETNSSLDPNDA--SEDPIL 578  
Db 1556 TIDSEKLLDDSDDDIEILECIISAMPKSRKKKPAQTAGKLPVPARKKSQLPVY 1615  
QY 579 K-IISPKYSGPASSSSLSFGG-----LYGHHHPILNNNSLSISN-----NSNNSN 624  
Db 1616 KILPSQNLQOPQKIVSFTPGDMPRVYCVES--TPINFTATSLSDLTTPESPPELLAGE 1673  
QY 625 SHRNGSNRSPHSASPML-----AAVAQGYSA-----GNSLLTSSSSSIQ 665  
Db 1674 GVRGGAQSGEFEXKDTIPTBERSIDEAQGKTSVTIPELDNKAEBGDIABEINGAMP 1733  
QY 666 K-----MMASNIQRQINEOSGOESLNGVNSDCSSNNGSSSIQYKKPSTISVAKIIGCT 719  
Db 1734 KGKSHKPPRVKINDVQVQOASASSAPNIKQLD-----GKXKKPTSPVKPIPON 1782  
QY 720 DTSRFGASPNLISQOHSAAHLTHQOQOQOQSLAQBALCKTRPRKGY-RAYDDSLVEA 778  
Db 1783 TEYTRVRKKNDSKNNLNAERVFSDNKO--SKQNLKNNSKDPNDKLPNNEDR----- 1833  
QY 779 VKAVORCEM---SVFRAGSYGVPH-----STLEKYVERHLMRPRKREPPQDPLV 827  
Db 1834 ---VRGSFAPDSFHHHTPIGTPTCYCFSRNDSLSLDPDDDDVLSR----- 1876  
QY 828 GLTGPANKLQDKLACGPHGSKLSNALKNQNNQAAAAAATAAAAAATPNGLKPLFE 887  
Db 1877 -----EXAEIRKAKENKESPAKVTSHTELTSSQNSANKTOAIA-----KOPTNR 1920  
QY 888 AGPOLSTQPMFPQTN-----ATNAVGLDPNRTITEMNR-POASNHGMLKSAOD 938  
Db 1921 GQPKRI-LQXOSTPPOSKDIPDRGAATDEKLQNF-----AIENTPVCFSHNSLSLSLD 1974  
QY 939 MVE---NYVDGIIIRKTLQASEGNGSAAAGNSGNGHGHGHGHALLDQLVKTPL 995  
Db 1975 IDQENNNKENPIKET--EPPOSGEP-----SKPOASGVARPSFH-----VEDTVP 2019  
QY 996 PFTNHRN-----NDYAATCSSAS-----GESVKGSGSPMG-----N 1026  
Db 2020 CFSRNSLSLSISIDEDLLQECISSAMPKKKPSRLKGDEKSPNMGCILGEDTLTD 2079  
QY 1027 YADIKR---ERLADS-----CGSSDEEH----- 1047  
Db 2080 LKDIORPDSEHGLPDSENFDWKAIOEGANSIVSLHAAAAAACLSSQASDSISLSLK 2139  
QY 1048 -----SASHINNNNSDLAHNKNKSGGGGGGNGQNTNGNRSSRMTSRDSEYDASSFK 1101  
Db 2140 SGISLGSFHLTPQOEKPTSNK-----GRIILKPGKSTLFTKXIESE----- 2185  
QY 1102 SGENGQGNHKKMDLNGSSSSSHIKCE-----SEAAIGHSPGHHTTS 1145  
Db 2186 KGKGGKVVYKSL-ITGKVRNSSEISGQMKPQLQANMPSISRGRTMIHIFGVRRNS 2240

## RESULT 9

US-08-449-731-2

Sequence 2, Application US/08449731  
Patent No. 6413727  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
ANAND, RAKESH  
CARLSON, MARY  
GRODEN, JOANNA  
HEDGE, PHILIP J.  
JOSLYN, GEOFF  
KINZLER, KENNETH  
MARKHAM, ALEXANDER F.  
NAKAMURA, YUSUKE  
THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti, LTD  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,731  
FILING DATE: 25-May-1995  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/289,548  
FILING DATE: 12-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.46943  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-08-449-731-2  
Query Match 3.8%; Score 229.5; DB 4; Length 2843;  
Best Local Similarity 19.3%; Pred. No. 1.6e-07;  
Matches 266; Conservative 179; Mismatches 510; Indels 421; Gaps 63;  
QY 2 HISSEISLIERVAECMGRRO--WKHYODKLTCSHLNTE---EQQPIA-----LIGS 48  
DB 1054 HIEBEL---KSEQRQSHNOSTTTPVYESTDDKHLKQPHFGQDECVSPRSGANGS 1110  
QY 49 EDEPQYHNS-SKEISQSNPNHCKTENHRLQOHNQSQLEEDSENNOTSHDSRTPTP 107  
DB 1111 ETRNGSNHGINQNSQS---LCQEDDYEDDKPTVYSERYSEEE---QHEEERPTN 1161  
QY 108 GA---TSRPSPEPEIDMRPSAKNFCVNGRLITVNAQCKLVAESAATITSSSTNSHH 164  
DB 1162 YSIKYNEXRHVDQPTDY-----SLKYATTDIPSSQKOSF 1195  
QY 165 QHDSQNSASLPHHISSSSSNNNSGNRAR---HIAAGARA---TPAATPANSL 217  
DB 1196 SFSKSSSGSSKTEHMSSESTSTPSSAKQONLHPSASGRSGQPPKAAITCVSSIN 1255  
QY 218 LKLTITRRAAKWTSMDSMAQLAQFSLADFNLI NSLASQOQOQOQOQOIA-----S 268

DB 1266 QETIQTYVEDTPICFSSRCSLSLSLSAEDEICGNTTOEADSANTLQIAETKGIKIGTRS 1315  
QY 269 AVPTTSEVSAALIPALKDTPSPSVDAPLDLSKPSNPSISGCVKSVRACATPTPSGR 328  
DB 1316 AEDP-VSEVPVAVSQPRTK-----SRLOGSSLSSE--SARIKAVFEPGCA 1358  
QY 329 RAYSEEDLSRALQDVVANKLDARKASQHEORS-ILDNRLLFKMKHD-----Q 376  
DB 1359 KSPSK-----SGAQTPKSPREHYQETPLMRSCTSVSLDSFESRSIASSVQ 1406  
QY 377 EODHDG-----DELEDSDDAEAVDSNASTVYPAEFARAQRLKLSHSEHNSDL 428  
DB 1407 SEPCGWSGIIISPSDLPS--PGQTMPSRSKTPPPPTQAOTK-REV---PKAKAPT 1459  
QY 429 GEDVDRGSPKGRHPRACNASANQAPASIPLDANVLH--TLMAAGIMAPKLDDETQT 486  
DB 1460 AEKRESG-PKQ---AAVNAAVQR--VQVLPDADTLHPATSTPDGFGSSSLSLSL 1511  
QY 487 VGDPIKGLLVANSGINNEGLNLTLASQENSNGNASILLQOQOQHQQHQQOQOQOQ 546  
DB 1512 DEPIQ-----KVELRIMPVQENDNGN-----ETESQPKESNENQEKEX 1555  
QY 547 HVAAYRHRLPKSE-----TPETNSGLDPNDA--SEDPIL 578  
DB 1556 TIDSEKDLDDSDDDIELEECIISAMPTKSRKPKPAOTASGLPPVARKPQLPQV 1615  
QY 579 K-IPFVKXSGPSSSLSPGC-----LVGGHHHPLNNNSLSISN-----NNHNSN 624  
DB 1616 KLPSPQNLQPKHVSFTPGDMPRVYVEG--TPINSTATSLDLTIESPPNLAEG 1673  
QY 625 SHRNGSNRSPHSASPMI-----AAVAQGYSA-----GNSLITSSSSSIQ 665  
DB 1674 GVRGAQSGEFEPKRTITPBERSTDEAQCKTSSVTIPELDNKAEEDDILAEICNSAMP 1733  
QY 666 K-----MMASTIQRQINEQSGQESLRGNVSDCSNNGSSSLGYCKPSTISVAKITG 719  
DB 1734 GKSHKPRPVYKIMQVOOASASSAPKKNOLD-----GKKKKPTSPVPPIQON 1782  
QY 720 DTSRFGASPNLLSQOHHSAHNLTHQOQOQOQSLQSAEALCKGRPKRGKY-RVYDRSLVEA 778  
DB 1783 TEYRTVRKADSKNNLAEVFPDNDK--SKKNLKNNSDPPNDKLPNNEDR----- 1833  
QY 779 VKAVQGEH---SVHRAGSYGVPH-----STLEKVKYERHILMRPKREPKQPDLY 827  
DB 1834 ---VRGSFARSDPHHYTPIGTGYCFGRNSLSLDDDDVDLSR----- 1876  
QY 828 GLTGPAKQLQDKLAKGPHGSKLSNALKNQNNQAAAAAATAATPGLKPLFE 887  
DB 1877 ---EKAELRKAKENKESBAKVTSHTELSNQOSANKTOAIA-----KQPINR 1920  
QY 888 AGPQALSFQPMKFWQTN-----ATNAYGLDPRNRTTEAMRN-POASNHGMLKMSAD 938  
DB 1921 GQPKPI-LQKOSTFPOSSKDIPIRGAATDEKLQNR---AIENPVCFSHNSLSLSLD 1974  
QY 939 MVE---NYVDGIIRTTQASEGNGSAGNGSNGNGHGHGHGHALLDQLVKTPTL 995  
DB 1975 IDQENNNENPEIKET-EPPOSQEP-----SKPOASGTAPKSFH-----VEDTFV 2019  
QY 996 PFTNHRN-----NDYAATCSSAS-----GESVYKSGSPMG-----N 1026  
DB 2020 CFSRNSLSLSLIDBEDDLQECISAMPKKKPRRLKGDNEKHSFRWNGHILGEDTLTD 2079  
QY 1027 YADIKR---ERLSADS-----GSSSDEEH----- 1047  
DB 2080 LKDIQRPDSHGLSPDSNFPWKAIQEGANSIVSSLHQAALAAACLSROASDSDSITLSL 2139  
QY 1048 ---SASHINNNSDILAHNKNKSGCGGCGGCGNGQOTNGNSRPTSRDSETPASSFX 1101  
DB 2140 SGISLGSFPHLTPDQEKPTSNK-----GRIILKPGKSTLETETKISSES----- 2185  
QY 1102 SGENGQONHKMMDLNGSSSSSHIKCE-----SEATGHHSPGHHTTS 1145  
DB 2186 KGIGKGVYKSL-ITGKVRNSNSEISQMKOPLQANMPISIRGRTHMIIHFGVRNS 2240







LENGTH: 2842 amino acids  
 TYPE: AMINO ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: APC  
 US-07-741-940-7

Query Match 3.7%; Score 221.5; DB 1; Length 2842;  
 Best Local Similarity 19.3%; Pred. No. 6,1e-07;  
 Matches 265; Conservative 179; Mismatches 511; Indels 421; Gaps 63;

```

QY 2 HISSVEISLIERVAEECMGRQ---WKHYODKLTCSHLNIE---EQQPIA-----IAGS 48
DB 1053 HIIIDEI---KQSEORQSRNOSTTTPVYTESTDDKHLKQPHFGQOECSVYRSRGANGS 1109
QY 49 EDEPQVYHNS-SKEISQSNPNHCKTENHRLQOHNGSOLLEEDSENNQTSRPTPTP 107
DB 1110 ETNRVGSNHGIGNQVNSQ---LCQEDDYEDDKPTVYSEVSEF-----QHHEBERPTN 1160
QY 108 GA---TSIPSPPEPIDMRPSAKNFCVNGRLTLVNAQGLVAESAATATSTSTNSHIH 164
DB 1161 YSIKYNEEKRHVDQPIDY-----SLKYATDIPSSQKQSF 1194
QY 165 QHDSQSNSSASLPHHISSSSSNNNSGGRAR---HIAAASARA---TPAATATANGSE 217
DB 1195 SFSKSSSQSSKSTHEMSSSENTSTPSSNAKRONOLHPSSAQSRGQPOKAATCKVSSIN 1254
QY 218 LYKLTORAAKWTSMDSMAAQLAOFSLADPFLINSLSAQOQQOQQOQIA-----S 268
DB 1255 QETITCYCEDTPICFSSRSSLSLSAEDICGCTOTQOEAASANTLQIAETKEKIGTRS 1314
QY 269 AVTPTTSEVSAASIPALKDTPTSPSVDAFLDISKSPNSISGDKVYACATPTPSGR 328
DB 1315 AEDP-VSEVPVAVSQHPRTK-----SSRLQSSLSLSE--SARHKAVEFFSGA 1357
QY 329 RAYSEDELSRALQDVANKLDARKASQHEQRS-ILDNRLFKMKHND-----Q 376
DB 1358 KSPSK-----SGAQTPKSPPEHYQETPLMFSRCTSVSLSLSEFSSRSIASSVQ 1405
QY 377 EODHDC-----DELESDNDAAEVDNSASTVVPAPFAQLRLKLSHEHNSGL 428
DB 1406 SEPFGMWSGIISPSDLPS--PGOTMPSPRSKTPPPPTQTK-REV---PKNKAFT 1458
QY 429 GEDVDGRSPKQGRHPACGNASANOGAPASIPLDANVLH--TLMLAAGIGAMPKLDQTQ 486
DB 1459 AEKRESG-PKQ---AAVVAAYQR---VQVLPDADTLHFAFATESTIPDGSSCSLSLSL 1510
QY 487 VQDFIKGLLVANSGGIMNEGILNLISASQENSGNASLLQQOQHQQHQQHQQOQQO 546
DB 1511 DEPTIQ-----KVELRLIMPVQENDNGN-----ETESQPKESNNOKEKEX 1554
QY 547 HVAAYRHLRPKSE-----TPETNSLDPNDA---SEDPIL 578
DB 1555 TIDSEKDLDDSDDDILEECITISAMPTKSRKAKPAQTAQKLPVPVAKPQQLPVY 1614
QY 579 K-IPSFKVSGPSASSSLSPG-----LVGHHHPLNNSNLSISN-----NSNHSN 624
DB 1615 KLPFGONRLQPOKHVSFTPGDMPRVYCVEG--TPIINSTATSLDLTIESPNTLAGE 1672
QY 625 SHRNSNSPMSASPMI-----AAVAQCGYSA-----GNSLITSSSSSIQ 665
DB 1673 GVRGAQSGEEFKDITPTEGSTDEAQGKTSVTIPELDONKAEGBGIIAECINSAMP 1732
QY 666 K-----WMAANIQRQINQSGQESLRNGNVSDCSSNNGSSSLYKKPSTISVAKITGT 719
DB 1733 KQKSHKPRVKKIMQVQAASASSAPKNQND-----GKKKKPTSPVKPIPON 1781
QY 720 DTSRFGASPNLSQOHSANHLTHOQOQOQSLAQEALGKTRPKRGKY-RNYDRDSLVEA 778

```

```

DB 1782 TEYTRVRKNAKSDKNLNAERVFSDNKD---SKQNLKNNSKDFNDKLPNNEDR----- 1832
QY 779 VKAVORGEM---SVYRAGSYGVPH-----STLEYKVKERHLMRPKREPKQPDLY 827
DB 1833 ---VRGSFAPDSPHHYTPITPTPCFSRNDLSLSDPDDDDVDLSR----- 1875
QY 828 GLTGPAKLOLDKLCAGPHGSKLSNALKNONQAAAAAATAATPGLKLPFE 887
DB 1876 -----EKAELRKAKENKESSEKVTSHTELTNSNOQSAKTAIA-----KQPINR 1919
QY 888 AGPOLASTQPNMFWPQTN-----ATNAYGLDFNRTITAMRN-POASNHHGLKMSAD 938
DB 1920 GQPKPI-LOKOSTFPOSSKDIIDRGATDEKLONF-----AIENTPVCFSHNSLSLSD 1973
QY 939 MVE---NYVDGIRKTLQASBNGSAANGNSGNGCHGHCHALLDOLLYKTPPL 995
DB 1974 IDQENNNKENEPIKET-BPPSQGEP-----SKPQASGAPKSFH-----VEDPVP 2018
QY 996 PFTNHRN-----NDYAATCSGAS-----GESVYKRSQSPMG-----N 1026
DB 2019 CFSRNSLSLSLSDSEDDLQECTISSAMPKKKKPRLKGDNEKHSPRMNGGILGEDLTLD 2078
QY 1027 YADIKR---ERLADS-----GSSDDEH----- 1047
DB 2079 LKDIDRPPSEHGLSPDSENPDMKAIQEGANSIVSLHQAIAAACLSROASSPSDSLSTK 2138
QY 1048 ---SASHNNNNSDLAHKKNKSGGGGGGNGQNTGNRSSMTSRDSETPASSFK 1101
DB 2139 SGISLGSFPPLTPDOEKPTPSNK-----GPRILKPEKSTLETETKIESES----- 2184
QY 1102 SCGENQOQNRKMDLNGSSSSSHIKCE-----SEAAATGHSFGHHTTS 1145
DB 2185 KGIGKGKVVYKSL-ITGKVRNSNSETSGMKQPLQANMPSISRGRTMHHIPGRNNS 2239

```

# RESULT 12 US-08-289-548A-7

Sequence 7, Application US/08289548A  
 Patent No. 5648212

## GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEORGE

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Allegretti, LTD

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/289,548A

FILING DATE: 12-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107,46943

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 3.7%; Score 221.5; DB 1; Length 2842;
Best Local Similarity 19.3%; Pred. No. 6,1e-07;
Matches 265; Conservative 179; Mismatches 511; Indels 421; Gaps 63;

QY 2 HISSEIIELEVAECMGRQ--WKHYQDKLTCSHLIE---EQQPIA-----IAGS 48
Db 1053 HIEDEI--KQSEBORGRNOSTTYVYTESTDHLKFOHPHGOECVSPYRSRGANGS 1109
QY 49 EDEPSQYNHS-SKEISQGNPNHCKTENHRLREQHNGSOLLEBEDSENNQTSHDSRTPTP 107
Db 1110 ETNRGSHGHNQNVSSQ--LCQEDDYEDDKPTMYSERYSEE-----QHEBERRTN 1160
QY 108 GA---TSTPPSPPEFIDMRPSAKNFCVNGRLLTVNAOGLVAESAATATSSSTNSHIH 164
Db 1161 YSIKYNEEKRHVDPIDY-----SLKATATIPSSQKQSF 1194
QY 165 QHDSNDSASALPHHISSSSSNNNSGNRRAR---HTAASARA---TPPAATANBLE 217
Db 1195 SPSRSSSQSSKTEHSSSSSENTSTPSSNAKQONQLHSSASQSRGQPOKATCKVSSIN 1254
QY 218 LYKLTQRAAKTMSDMSAAQLAFSLADFLINSLASQOQOQOQOIA-----S 268
Db 1255 QETIQTVEEDTPICFGRCSLSLSADELICQNTTOEADSNATLQIAIEKEKIGTRS 1314
QY 269 AVPTTSVSAASAIIPALKDTPSPSVDAPLDLSKPSNNSISGDKVSRACATPTPSGR 328
Db 1315 AEDP-VSEVPVAVSOHPRTK-----SSRLGSSSISSE--SARHKAVERFSSGA 1357
QY 329 RAYSEEDLSRALQDVVANKLDARKSAQHEQRS--ILNRLFKMHND-----O 376
Db 1358 KSPSK-----SGAOTPKSPPHVVOETPLFMSRCTSVSLDSFESRSIASSVQ 1405
QY 377 EQDHG-----DELESDNDAAEAVDSNASTFVPAEFARQRLKLSLSEHNGSDL 428
Db 1406 SEPCSGWVGIIISPDLPS--PGQTPSPRSKTPPPPPQTAQTK-REV---PKNKAPT 1458
QY 429 GEDVDKRGPKMGRPHACGNASANOGAPASIPLDANVLH--TLMAAGIGAMPKIDETQT 486
Db 1459 AEKESG-PKQ---AANAAVAVOR---VQVLPDADATLLHFATSTPDPFSSSSLSLSL 1510
QY 487 VGDPIKGLLVANSGSINNEGILNLISASQENSNGASLLQQCQHQHQHQHQHQHQHQ 546
Db 1511 DEPIQ-----KVELRLMPVOENDNGN---ETESQPRESENDEKKEK 1554
QY 547 HVAAYRHLRKPSE-----TPETNSSLDPNDA---SEDPIL 578
Db 1555 TIDSEKDLDDSDDDDIILEECISAMPTKSSRAKAPQATASLTLPVPARKPSQLP 1614
QY 579 K-IPGFVSGPSSSSSLSPG-----LVGHHHPLNNNSLSISN-----NSNHSN 624
Db 1615 KLPSONRLOPQKHSFTPGDDMPRYCVEG--TPINFSTATSLDLTIESPNIAGE 1672
QY 625 SHRNSNNSPHSASML-----AAVAAGGYSV-----GNSLTSSSSSIQ 665
Db 1673 GVRGAQSGEERKDJITEGRSTDEAGGKTSVTIPELDNKAEBEDDIAECINSAMP 1732
QY 666 K-----MMAAGNIQROIENGQESLRNGNVSDCSNNGSSSLGKPKPSIAKIGCT 719

```

```

Db 1733 KKGSHKPRVKKINDQVQASASSAPNKQLD-----GKXKPSVPKPIQN 1781
QY 720 DTSRFGASPNLSQOHHSAHHLTHQOQOQOQLSAQBALCKGTRPKRGKY-RNYDPSLVEA 778
Db 1782 TEYRTRVAKNADSKNNAEAEVFPDNDK---SKQONKNNSKNDPNDKLPNNEDR----- 1832
QY 779 VKAVQRGEM---SVHRASYYGVH-----STLEKHYERHLMRKRKEPRQPDV 827
Db 1833 ---VRGSFAPDSPHHTPIEGTPTCYCFSRNDSLSLDPDDDDVDLSR----- 1875
QY 828 GLTGPANKLQDLKAKGPHGSKLSNALKNQNNQAAAAAATAATNGUKLPLE 887
Db 1876 -----EKAEELRKAKENKESPAKVTSHTELTNSQSANKTQALA-----KQPINR 1919
QY 888 AGPOLSFQPMFPPQTN-----ATNAYGLDPNRTITEMMRN-POASHHGLMKSD 938
Db 1920 QQPKPI-LQKQSTPQSSKQIDPRGAATDEKIQNF-----AIENTPVCFSHNSLSLSLD 1973
QY 939 MVE---NYVDGIIKRTLOASEGNGSAAAGNSGNSGNGHGHGHALDLQLLVKTPL 995
Db 1974 IDQNNKNKEPIKET-EPDPSQEP-----SKPQASGTAAPKSFH-----VEDTIV 2018
QY 996 PFTNHRN-----NDYAATCSSAS-----GESVKRSGSPMG-----N 1026
Db 2019 CFSRNSLSLSLSISEDDLLQECISAMPKKKPSRLKQDNKKSPPNMGGILGEDTLTD 2078
QY 1027 YADIKR---ERLADS-----GGSDDEH----- 1047
Db 2079 LKDIORPSEHGLSPDSENPFWKAIQEGANSIVSLQAAAAACLSRQASDSDSISLTK 2138
QY 1048 ---SASHNNNSDLAHKNKSGGGGGGGGQNGQNTNGNRSRMTSDSETPDASSFK 1101
Db 2139 SGISLSPFHLTPQOEKPTSNK-----GRILLKPKGKSTLETYKLESSES----- 2184
QY 1102 SGENGQGNHMKMDLINGSSSSSHIKCE-----SEATGHSFGHHTTS 1145
Db 2185 KGIGKGKVVYKSL-ITGKVRNSNSEISGQMQLQANMPSISRGRTWHIPEVRNRS 2239

RESULT 13
US-08-452-654-7
Sequence 7, Application US/08452654
Patent No. 5691454
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995

```

```

: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/741,940
: FILING DATE: 08-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Kagan, Sarah A.
: REGISTRATION NUMBER: 32,141
: REFERENCE/DOCKET NUMBER: 1107.035574
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-508-9100
: TELEFAX: 202-508-9299
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2842 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: IMMEDIATE SOURCE:
: CLONE: APC
: US-08-452-654-7

```

Db	1615	KLUSQNRLODQKXIVSTTPGDDMRVYCVES--TPINFTASTSDLTIESPPELLAAGE	1672
Oy	625	SHRNGSNRSPHSASPML-----AAAVAOGGYSA-----GNSLITSSSSSIQ	665
Db	1673	GVRGGAQSGEPKEDTITPEGRSTDEAOGGKTSVITPELDNKAEEGDILAEICINSAMP	1732
Oy	666	K-----MMASNITORINEOSGGSSLLNGVNSDCSSNNGSSSLGYKKPISIVAKIIGCT	719
Db	1733	KGKSHKPRVYKIMDVQOQASASSAPKXQLD-----GKKKKPTSPVKPIPON	1781
Oy	720	DTSRFGASPMLSQOCHSAHNLITQOQOQOOLSAOEALGKGRPRKGY--RYDYRDSLVEA	778
Db	1782	TEYTRRKRKADSKONLMAERVFDNND---SKONLKNSKDNKDLPNNEBR-----	1832
Oy	779	VKAVORGEM--SVHRAAGSYGVPH-----STLEYKYKERHLMRPRKPRQPDV	827
Db	1833	---VRGSFAPDSPHHNTPIEGTGYCFESRNDLSLDFDDDDVDSL-----	1875
Oy	828	GLTPRAKLTQLDKACGRHGSKLSMNLKONONAAAAAATAATNGKLPFE	887
Db	1876	---EVAELTRKAKKENKESEAKVTSTELTSSNOOSANKTOALA-----KOPINR	1919
Oy	888	AGPOALSFQPMFMPQTN-----ATNAYGLDFNRPITEAMRN--POASNHGKMSKAD	938
Db	1920	QGPKPITLQKQSTPQSSKQIDPBGATDELYQNF---AIENTPUCFSINSSLSLSD	1973
Oy	939	MVE---NVYDGIIRKTLQASEGNGSAAAGNSGNSNGHGHGHGHALLDOLLYKTP	995
Db	1974	IDQENNNKENEPKIKET--EPDPSQGER---SKPOASGYAPKSPFH-----VEDP	2018
Oy	996	PFTMHRN-----NDYATCSAS-----GESVKRSGSPMG-----N	1026
Db	2019	CFSSNSSLSTLSDISEDDLQECISSAMPKKKKPRSLKGDNEKKS PRNMGGILAEBLTLD	2078
Oy	1027	YADIKR-----ERLSADS-----GSSSDEEH-----	1047
Db	2079	LKQIDORDSEHGSLPDBENPFWKALIOGCAIVSALHQAAMAAALSRQASDSBSILSK	2138
Oy	1048	-----SASHINNNSDLAHNKKNKSGGGGGGGGQTNNGRSGRMTSRDSETDASSFK	1101
Db	2139	SGISLGRPFHLTPQOEKPRFTSNK-----GRILKPGKSTLETETKIESES-----	2184
Oy	1102	SGENGGOONHMDLNGSSSSSHIKCE-----SEATGHNSPCHNHTS	1145
Db	2185	KGIGKGGKKVYKSL--ITGKVSNSNBSICQOKOPLOANPISIRGRTMIIPEVRNNS	2239

RESULT 14  
 US-08-449-731-7  
 ; Sequence 7, Application US/08449731  
 ; Patent No. 6413727  
 GENERAL INFORMATION:  
 APPLICANT: ALBERTSEN, HANS  
 ANAND, RAKESH  
 CARLSON, MARY  
 GRODEN, JOANNA  
 HEDGE, PHILIP J.  
 JOSLYN, GEORGE  
 KINZLER, KENNETH  
 MARKHAM, ALEXANDER F.  
 NAKAMURA, YUSUKE  
 THILLYERIS, ANDREW  
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 GENE IN COLORECTAL CANCER IN HUMANS  
 NUMBER OF SEQUENCES: 102  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Banner & Allegretti, LTD  
 STREET: 1001 G Street, NW  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20001-4598

```

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,731
FILING DATE: 25-May-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,548
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9299
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-449-731-7

Query Match      3 7%: Score 221.5; DB: 4; Length 2842;
Best Local Similarity 19.3%; Pred. No. 6,11e-07;
Matches 265; Conservative 179; Mismatches 511; Indels 421; Gaps 63;

QY 2 HISSYEILSERVAEECMGRQ---WKHYODKLTCSHNLTE---EQOPIA-----IAGS 48
DB 1053 HIIIEI---KQSEBORQSRNSTYPTVTESTDDHLKQFHPHQEQECVSPRSKANGS 1109
QY 49 EDEPSQYHNS-SKEISQSNPNKCTENHRLLEQOHNGSQLLEEDEENNQTSHDSRTPP 107
DB 1110 ETRNGVSNHGINONVQS---LCQEDYEDDKPTNYSRYSBEE-----QHEEERPTN 1160
QY 108 GA---TSTPPEPPIIDWRPSAKCNFCVNGRLTVAAGKLVAAEATATSSSTNSIHH 164
DB 1161 YSIKYNEEKRHVDQIDY-----SLKYATDIPSSQKQSF 1194
QY 165 QHSDSNSSASLPHHIISSSSSSNNNSGNRAR---HIAASARA---TPAATPANGLE 217
DB 1195 SFSKSSSGQSKTEHMSSESTSTPSSNAKQONLHSSAOSRSGQKKAATCVSSIN 1254
QY 218 LYKLITQRAAKTMSNDMAAQLAOFSLADENLINSLASQOQOQOQOIA-----S 268
DB 1255 QETITQYCEVEDTPICFSSCSLSLSAEDIGCQTTQZADSANTLQIAETIKETGRS 1314
QY 269 AVPTPTSEVSAALSPALKDTPSPVDAPLDLSKSPSSNSISGDKVSKVACATPTPSGR 328
DB 1315 AEDP-VSEVPVAVSQHPTK-----SSRLQSSSLSSSE--SARHKAVERFSSGA 1357
QY 329 RAYSEEDLSRALQDVANKLDARKASQHEQRS-ILDNRLFKMGHND-----Q 376
DB 1358 KSPSK-----SGAQTPKSPRPHYQETPLMFSRCTSVSLDSFESRSIASSVQ 1405
QY 377 EQDHG-----DELEDSNDAAEVDNASTPVVPAEPARAQLRLKLSLSEHNGSDL 428
DB 1406 SEPSCGWSGIIISPDLPDS--PGQTPMPSRSKTPPPPPQATK-REV---PKKAPPT 1458
QY 429 GEDVDKSPKGRHACGNASANOAPASIPLDANVLH--TLMAGIGAMPKLDDETQOT 486
DB 1459 AKRESG-PKO---AAVNAVAOR---VOVLDADDTLLHFTASTPTDFGSCSSSLALSL 1510

```

```

QY 487 VGDFIKGLLVANSGGIMNEGILNLNLSAQSNSNGASILLQOQOHOQHQQOQOQOQ 546
DB 1511 DEPIQ-----KDVLRIMPVQENDN-----ETSEQPRESENENEKEXEK 1554
QY 547 HVAARHRLPKSE-----TPTNSSLDPNDA---SEDPIL 578
DB 1555 TIDSEKLLDSDDDIEILECTISAMPKSRKAKKPAQTASKLPPVARKPSQLPVY 1614
QY 579 K-ISFKVSGPASSSSISPG-----LVGHHHPLNNNSLSLN-----NSNHSN 624
DB 1615 KILPSQNLQPKQKVSFTPGDMPRVYCVES--TPINSTATSLDLTITESPNEBLAGE 1672
QY 625 SHRNGSNRSPHSASPML-----AAVAQGYSA-----GNSLLTSSSSSIQ 665
DB 1673 GVRGAQSGEEKEDTITPEGRSTDEAOGKTSVTTPELDONKABESDILAECINSMP 1732
QY 666 K-----MMASNIQROINEQSGESLRNGVSDCSNNGSSSIQYKPSISVAKITGCT 719
DB 1733 KKGSKHPPRVKINDVOQASASSASPKNQOLD-----GKKKKKPSPYKPIPON 1781
QY 720 DTSRFGASPNLISQOHSANHLTHQOQOQOQSLAQOALGKGRPRKGY-RVYDRDSLVEA 778
DB 1782 TETRTVRKNDSPKNNLNAERVFSDNKO--SKQNLKNSKDPNDKLPNNEDR----- 1832
QY 779 VKAVQSGEM---SVHRAGSYYGVPH-----STLEKVKERHMRPRKKEPKQPQDLV 827
DB 1833 ---VRGSFAPDSPHHTPIEGTPYCFSRNDSLSLDFDDDDVLSR----- 1875
QY 828 GLTGEPANKLQDLKLAGPHGSKLSNALKNONNAAAAAATAATPNGIKLPLFE 887
DB 1876 ---EKAEELRKAKENKSEAKVTSHTELSNQSANKTOAIA-----KQPINR 1919
QY 888 AGPOLSPQPMFAPQTN-----ATNAYGLDFNRTIETAMR-POASNHGMLKMSQD 938
DB 1920 GQPKPI-LQKSTPQSSKIDIPDGAATDEKLQNF---AIEVTPCFSNSSLSISD 1973
QY 939 MVE---NVYDGIIRKTLQASEGNSAANGSNGNGHGHGHALLDQLLVKKTPL 995
DB 1974 IDQENNNKENEPIKET-EPPSQGEF-----SKPQASGVAPKSPH-----VEDTPI 2018
QY 996 PFTNHRN-----NDYAATCSSAS-----GESVKRSGSBMG-----N 1026
DB 2019 CFSRNSLSLSLSIDSEDLLOECISAMPKKKPSRLKGDMEKSPRMGGIILGEDTLTD 2078
QY 1027 YADIKR---ERLADS-----GGSSDEH----- 1047
DB 2079 LKDIQRPDSERGLPDSBNFDWKAIQGANSYVSLHQAALACLSSQASDSISLSLK 2138
QY 1048 ---SASHINNNSDLAHNKNKSGGGGGGNGQNTNGRSSRMTSRDSETDASSFK 1101
DB 2139 SGISLSPFHILTPQEEKPFTSNK-----GRIILKPGKSLLETKTIESES----- 2184
QY 1102 SGENGGOONHMDMLNGSSSSSHIKCE-----SEATGHSPPHHTTS 1145
DB 2185 KGIGGKKVYKSL-ITGKVRNSSEISGOMKOPLOANNPISIRGRTHIIPGVNRS 2239

RESULT 15
US-08-452-655B-2
; Sequence 2, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

```

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,655B  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,548  
FILING DATE: 12-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.49964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-452-655B-2

Query Match 3.7%; Score 221.5; DB 1; Length 2843;

Best Local Similarity 19.3%; Pred. No. 6,1e-07;

Matches 265; Conservative 179; Mismatches 511; Indels 421; Gaps 63;

QY 2 HISSEYELSERVAEECMGRRO--WKHYQDKTCSHLNIE--EQQPIA-----IAGS 48  
DB 1064 HIEDEI--KQSEOROSRNOSTTYPTTESTDDKLFQPHFGQCEVSPYRSGANGS 1110  
QY 49 EDEPSQVYNS-SKEISQGNPNHCKTENHLEQOHNGSOLLEEDSENNQTSHTSRTPTP 107  
DB 1111 ETNRVSGNHGINQNVSGS---LQOEDYEDDKPTMYSRYSSEF-----QHEEERTN 1161  
QY 108 GA---TSPTSPPEPPIIDWRPSAKNFCVNGRLLITVNAQGLVAESAATATSSSTNSHIH 164  
DB 1162 YSIKYNEEKRHVDQPIDY-----SLKYATDIPSSQKQSF 1195  
QY 165 QHDSNNSASLPHHSSSSSSNNNSSGNRAR---HIAASARA---TPAATRANLE 217  
DB 1196 SPSKSSSSQSSKTEHMSSESTSTPSSNAKRONOLHPSASQSRGQPOKATCKVSSIN 1255  
QY 218 LYKLITQRAAKWTSMDSMAAQLQPSLADFNILNSLASOQQOQQOQOIA-----S 268  
DB 1256 QETIOTYCEVDPTPICFSRCSLSLSASADEIGCQOTQOADSANTLQIAETKEKIGTRS 1315  
QY 269 AVPTTTSVSAALISPALKDTPPSPVDAPLDSKSPSPNSGISGVKSVACATPTPGR 328  
DB 1316 AEDP-VSEVPVAVSQHPRTK-----SSRLQSSLSSE--SARHKAVERFSSGA 1358  
QY 329 RAYSEEDLSRALQDVAVAKLDAKRSASQHEQRS--ILDNRLEFKMGHND-----Q 376  
DB 1359 KSPSK-----SGAQTPKSPPEHYVQETPLMFSRCTSVSSLDSFESRSIASSVQ 1406  
QY 377 EODHDG-----DELEDSNDDAEAEVDSNASTPVYPAFAQLRKLKLSLSEHNGSDL 428

DB 1407 SEPCGWNVSGLIISDLPDS--PGQTMPSRSKTPPEPPOTAQTK-REV----PKNKAPT 1459  
QY 429 GEDVDGRSPKMGRRPACGNASANOAPASIFLDANVLH--TLMLAIGAMPKLDERTQ 486  
DB 1460 AEKESG-PKO---AANAAVOR---VOVLPDADTLHFPTSTPPDFSCSSISALS 1511  
QY 487 VGDFIKGLLVANSGGINNEGLNLNLASQENSNGNASLLQOQOHOQHOOHQOQOQO 546  
DB 1512 DEPIQ-----KDELRLIMPVQENDNG--ETESQPKESNENQEKAEK 1555  
QY 547 HVAAYRHRPKSE-----TPETNSLDPNDA--SEPI 578  
DB 1556 TIDSEKLLDSDDDIELEECIISAMPTRSKAKKAPQASTKLPPVARKSOLPVY 1615  
QY 579 K-IPSFVSGPSSSSISPGG-----LVGHHHPILNNNSLSISN-----NSNHSN 624  
DB 1616 KLPSQNRLOPKHVSFTPGDMPRVVCVEG--TPINSTATSLDLTITESPPNELAGE 1673  
QY 625 SHRNGNRSPPHSASPMI-----AAVAQGYSA-----GNSLLTSSSSSIQ 665  
DB 1674 GVRGAQSGEPKEDTITPTEGRSTDEAGGKTSSVTIPEDDNKAEBDILAEICNSAMP 1733  
QY 666 K-----MNASNIOQINEQGESLRGNVSDCSSNNGSSSLGKYPKPSISVAKIIGT 719  
DB 1734 KGSKHKPRVKKIMDOVOQASASSAPYKQOLD-----GKKKKPTSPVPIPN 1782  
QY 720 DTSRFGASPNLSQOHSANHLTHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 778  
DB 1783 TEYRTRVRKKNADSKNNLAERVFSDNKO--SKQNLNKNKDPNDKLPNNEDR----- 1833  
QY 779 VKAIVORGE--SVRAQSYGVPH-----STLEKVKERHLMRPRKEPKQPDLY 827  
DB 1834 ---VRGSFAPDSHHYPIIGTPYCFBRNLSLSDPDDDDVDLSR----- 1876  
QY 828 GLTGPANKLQDKLACGPHGSKLSNALKNONQAAAAAATAATNGLPLFE 887  
DB 1877 ---EKAEIRKAKENKESAKVTSHETLSNOQSANKTQAI-----KQPINR 1920  
QY 888 AGPQALSQPMFMFQTN-----ATNAAGLDPNRTTEAMRN--PQASNHGMLKSNOD 938  
DB 1921 GQPKPI-LQKOSTPQSSKDIIDRGAATDEKLNQF--AIENTPVCFSHNSLSLSD 1974  
QY 939 MVE---NVYDGIIRKTLQASEGNSAAGNSGNSGNGHGHGHALLDQLLVKKTPL 995  
DB 1975 IDQENNKENPIKET--EPPOSQEP-----SKPQASGYAPKSFH-----VEDTPV 2019  
QY 996 PTNHRN-----NDYAATCSSAS-----GESVKRSGSPWG-----N 1026  
DB 2020 CFSRNSLSLSISIDSEDDLQECISAMPKRRKSRLLKGNDEKISPRMGGILGEDTLTD 2079  
QY 1027 YADIKR---ERLSADS-----GSSSDEEH----- 1047  
DB 2080 LKDIORPSEHGLSPDSENPDMKAIQEGANSIVSLHQAALAAACLSROASDSISLSLK 2139  
QY 1048 ---SASHINNNSDIAHNKNSGGGCGGNGQNTGNRSMRSDSETPASSFX 1101  
DB 2140 SCISLGSFHLTPPOEKFPKTSNK-----GRILKPGKSTLETUKKISES----- 2185  
QY 1102 SGENGQGNHMKMDLNGSSSSSHIKCE-----SEATGHSFGHHTTS 1145  
DB 2186 KGIRGKKVYKSL-ITGKVRNSSEISGQMKPLQANMPSIRGRTHIHPVRNNS 2240

Search completed: October 29, 2003, 12:32:08  
Job time : 45 secs

THIS PAGE BLANK (USP)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Gen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 12:21:56 / Search time 28 Seconds

(without alignments)  
1956.647 Million cell updates/sec

Title: US-10-016-768a-10

Perfect score: 6030  
Sequence: 1 MISSISYEISLSEVABECMGR.....ILHEKLAQIKAEQVDADQL 1165

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	311.5	5.2	2038	1 FSH_DROME	P13709 drosophila
2	299.5	5.0	2175	1 HMCU_DROME	P10180 drosophila
3	285.5	4.7	1211	1 BUN2_DROME	Q24523 drosophila
4	272.5	4.5	1073	1 HR38_DROME	P49689 drosophila
5	272.5	4.5	1669	1 ASX_DROME	Q9V727 drosophila
6	271	4.5	1365	1 SUZ2_DROME	P25172 drosophila
7	259	4.3	1028	1 OVO_DROME	P51521 drosophila
8	257.5	4.3	1077	1 HLES_DROME	Q02308 drosophila
9	249.5	4.1	1556	1 PROS_DROVI	Q09641 drosophila
10	249	4.1	1165	1 YNF4_YEAST	P53950 saccharomyc
11	246.5	4.1	880	1 BRCA4_DROME	Q24206 drosophila
12	244.5	4.1	883	1 E7AB_DROME	P11536 drosophila
13	243	4.0	1507	1 SIVA_DROME	Q24167 drosophila
14	238.5	4.0	1403	1 CYAL_DROME	P32870 drosophila
15	237	3.9	2248	1 AF17_HUMAN	P55198 homo sapien
16	231	3.8	1093	1 AP17_HUMAN	P17672 drosophila
17	231	3.8	1394	1 E75B_DROME	P45447 drosophila
18	230.5	3.8	865	1 E75B_DROME	P45447 drosophila
19	228	3.8	1060	1 ZFH1_DROME	P28166 drosophila
20	225.5	3.7	1983	1 TF20_MOUSE	Q09648 mus musculu
21	223	3.7	727	1 BRCA1_DROME	Q09648 mus musculu
22	223	3.7	1960	1 TF20_MOUSE	Q09648 mus musculu
23	222	3.7	1253	1 DSRP_HUMAN	Q9UWU0 homo sapien
24	221.5	3.7	2843	1 APC_HUMAN	P25054 homo sapien
25	221	3.7	1230	1 ST20_CANAL	Q92212 candida alb
26	220.5	3.7	448	1 AAC2_DICDI	P14196 dictyostell
27	220	3.6	960	1 YMX6_YEAST	Q04279 saccharomyc
28	220	3.6	989	1 PTPX_DICDI	P54637 dictyostell
29	219.5	3.6	1140	1 YMX6_YEAST	Q04893 saccharomyc
30	216.5	3.6	635	1 HMLA_DROME	P10105 drosophila
31	216.5	3.6	1403	1 CTC_DROME	Q9U105 drosophila
32	215	3.6	758	1 YMX8_YEAST	Q03825 saccharomyc
33	214	3.5	1596	1 MAM_DROME	P21519 drosophila

34	212.5	3.5	1317	1 GAP_CAEEL	P34288 caenorhabdi
35	211.5	3.5	1237	1 E75A_DROME	P17671 drosophila
36	211	3.5	918	1 YMXB_CAEEL	P34487 caenorhabdi
37	211	3.5	1443	1 E75C_DROME	P13055 drosophila
38	210.5	3.5	1780	1 YK26_CAEEL	P34333 caenorhabdi
39	209	3.5	1043	1 FTFL_DROME	P33244 drosophila
40	209	3.5	2845	1 APC_MOUSE	Q61315 mus musculu
41	208.5	3.5	1411	1 TCOF_HUMAN	Q13428 homo sapien
42	208	3.4	1597	1 SOL_DROME	P27398 drosophila
43	206.5	3.4	931	1 LUG_DROME	Q9FUY2 arabidopsis
44	206	3.4	1319	1 MNI_HUMAN	Q10571 homo sapien
45	205	3.4	805	1 E2F_DROME	Q27368 drosophila

## ALIGNMENTS

RESULT 1  
ID FSH\_DROME STANDARD: PRT: 2038 AA.  
AC P13709: P13710:  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Female sterile homeotic protein (Fragile-chorion membrane protein).  
GN PS(1)H OR FSH.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89276730; PubMed=2567251;  
RA Haynes S.R., Mozer B.A., Bhatia-Dey N., David I.B.;  
RT "The Drosophila fish locus, a maternal effect homeotic gene, encodes  
RT apparent membrane proteins."  
RL Dev. Biol. 134:246-257(1989).  
CC -!- FUNCTION: REQUIRED MATERNALLY FOR PROPER EXPRESSION OF OTHER  
CC HOMOTIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.  
CC -!- SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.  
CC -!- SIMILARITY: Contains 2 bromodomains.  
CC -!- SIMILARITY: Contains 1 ET domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M23221; AAA28540.1; -  
CC EMBL: M23222; AAA28541.1; ALT\_TERM.  
CC EMBL: M15762; AAA70424.1; -  
CC EMBL: M15763; AAA70423.1; -  
CC EMBL: M15764; AAA70422.1; -  
CC FIR: A43742; A43742.  
CC FIR: A43742; A43742.  
CC HSP: Q92831; 1B91.  
CC FlyBase: FBgn004656; fs(1)h.  
CC InterPro: IPR001487; Bromodomain.  
CC Pfam: PF00439; bromodomain; 2.  
CC PRINTS: PR00503; bromodomain.  
CC SMART: SM00297; BROMO; 2.  
CC PROSITE: PS00633; BROMODOMAIN 1; 2.  
CC PROSITE: PS50014; BROMODOMAIN 2; 2.  
CC Developmental protein: Bromodomain; Transmembrane; Repeat.  
FT DOMAIN 51 123 BROMODOMAIN 1.  
FT DOMAIN 495 567 BROMODOMAIN 2.  
FT DOMAIN 945 1106 ET DOMAIN.  
FT TRANSMEM 330 350 POTENTIAL.  
FT TRANSMEM 451 471 POTENTIAL.  
FT TRANSMEM 750 770 POTENTIAL.







```

Db      1093 LIARPRETAPSPFLFSPSLFGAAGMGASNA-----PAMADEMNRHVEFEI 1143
Qy      995 LPTFNRRNDVAA-----TSSAGSEGVKSGSGMNYA 1028
Db      1144 AKLOOHQOOQAAQAQAFPFSSLMALQOOVLNGAODLSIAAAAAKDKILNG----- 1196
Qy      1029 DIKERLSADSGSGS---DEHSASHINNNSDLAHNNKSGCG-----GGGGGNG 1076
Db      1197 --QSSSLHSGSSSCSKDGERDDAYPS-----LHGKKSREGGTPAPAPPSGPGTAG 1249
Qy      1077 -----QTNG-NGRSMRSTRDSETPDASFSGSGENGQOONHKMDLNGSSSSSHIKC 1128
Db      1250 APPTAPPTGAGSSNSAPSPLSNLTLPALSSGGEFFAATAPSPQRN-----ASTNLI 1305
Qy      1129 ESEATGHSFCHHTTSLIHEKLAQIKAEVDQADQL 1165
Db      1306 TOPPVTPHSTPQRPTRKAV---LPITQOQDFMNNL 1339

RESULT 3
BUN2_DROME          STANDARD;          PRT; 1211 AA.
ID      BUN2_DROME
AC      Q24523; Q9VK78; Q9VK79;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Bunched protein, class 2/class 3 isoforms (shortsighted protein).
GN      BUN OR SHS OR CGS461.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila..
CX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORM CLASS 2), SUBCELLULAR LOCATION, AND
RP      FUNCTION.
RC      TISSUE=EYE-antennal disk; PubMed=7555710;
RX      MEDLINE=96038094; PubMed=7555710;
RT      "Shortsighted acts in the decapentaplegic pathway in Drosophila eye
RL      development and has homology to a mouse TGF-beta-responsive gene.";
RL      Development 121:2835-2845(1995).
RN      [2]
RP      SEQUENCE FROM N.A. (ISOFORMS CLASS 2 AND CLASS 3).
RC      STRAIN=Berkelley;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S.J., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.H.C., Blake J.R.G., Champagne M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abail J.F., Agbayani A., An H.-J., Andrews-Ffiankoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
RA      Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA      Buttis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA      Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA      Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jajaeli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Metcalf B., McIntosh N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pauleb J.M.,

```

```

RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA      Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195(2000).
CC      -!- FUNCTION: Probable transcription factor required for peripheral
CC      nervous system morphogenesis, eye development and oogenesis. May
CC      be required for the transmission of the dpp signal and for a
CC      morphogenetic movement of the medulla in the brain that reorients
CC      the second optic lobe relative to the first. Plays a role in
CC      determining proper dorsal cell fates leading to the formation of
CC      the dorsal appendages.
CC      -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC      -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=3;
CC      Comment=Experimental confirmation may be lacking for some
CC      isoforms;
CC      Name=Class 2;
CC      IsoId=Q24523-1; Sequence=Displayed;
CC      Name=Class 1;
CC      IsoId=Q24522-1; Sequence=External;
CC      Name=Class 3;
CC      IsoId=Q24523-2; Sequence=VSP_006670;
CC      -!- SIMILARITY: BELONGS TO THE TSC-22/DIP/BUN FAMILY.
CC      -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC      gene model prediction.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; LA2512; AAC41608.1; -; ALT_SEQ.
DR      EMBL; AE003636; AAF53200.1; ALT_SEQ.
DR      EMBL; AE003636; AAF53201.1; ALT_SEQ.
DR      PIR; T13804; T13804.
DR      HSSP; P80220; 1DIP.
DR      FLYBase; FBgn0010460; bun.
DR      InterPro; IPR000580; TSC-22_Dip_Bun.
DR      Pfam; PF01166; TSC22; 1.
DR      ProDom; PD007152; TSC-22_Dip_Bun; 1.
DR      PROSITE; PS01289; TSC22; 1.
KW      Transcription regulation; Nuclear protein; Alternative splicing.
FT      DOMAIN 15
FT      DOMAIN 31
FT      DOMAIN 76
FT      DOMAIN 97
FT      DOMAIN 102
FT      DOMAIN 237
FT      DOMAIN 241
FT      DOMAIN 249
FT      DOMAIN 254
FT      DOMAIN 261
FT      DOMAIN 265
FT      DOMAIN 306
FT      DOMAIN 321
FT      DOMAIN 322
FT      DOMAIN 328
FT      DOMAIN 348
FT      DOMAIN 356
FT      DOMAIN 607
FT      DOMAIN 619
FT      DOMAIN 661
FT      DOMAIN 668
FT      DOMAIN 743
FT      DOMAIN 746
FT      DOMAIN 759
FT      DOMAIN 765
FT      DOMAIN 795
FT      DOMAIN 801
FT      DOMAIN 817
FT      DOMAIN 822
FT      DOMAIN 832
FT      DOMAIN 838
FT      DOMAIN 884
FT      DOMAIN 891
FT      DOMAIN 927
FT      DOMAIN 947
FT      DOMAIN 1001
FT      DOMAIN 1005
FT      DOMAIN 1011
FT      DOMAIN 1014

```



RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.R., Kechum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulip D., Lai Z.,  
RA Laeko P., Lei Y., Lavitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mekulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclik J.M.,  
RA Palazzolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Silden-Klamas I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye Y., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT Science 287:2185-2195(2000).  
RL [5]  
RN SEQUENCE OF 528-1073 FROM N.A.  
RP STRAIN=Canton-S;  
RC MEDLINE=96068664; PubMed=7479849;  
RX Fisk G.J., Thummel C.S.;  
RA "Isolation, regulation, and DNA-binding properties of three  
RT Drosophila nuclear hormone receptor superfamily members".  
RL Proc. Natl. Acad. Sci. U.S.A. 92:10604-10608(1995).  
CC -I- FUNCTION: BINDS TO NGFI-B RESPONSE ELEMENTS. PLAYS AN IMPORTANT  
CC ROLE IN LATE STAGES OF EPIDERMAL METAMORPHOSIS.  
CC -I- SUBUNIT: FORMS A HETERODIMER WITH USP.  
CC -I- SUBCELLULAR LOCATION: Nuclear.  
CC -I- ALTERNATIVE PRODUCTS:  
CC Name=Long;  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=short;  
CC IsoId=P49869-1; Sequence=Displayed;  
CC Name=short;  
CC IsoId=P49869-2; Sequence=VSP\_003714;  
CC -I- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN PREBLASTODERM  
CC EMBRYOS, SPECIFICALLY IN CENTRAL NERVOUS SYSTEM AND INTESTINAL  
CC TRACT. HIGHLY EXPRESSED IN THIRD INSTAR LARVAL IMAGINAL DISKS AND  
CC BRAIN COMPLEXES, BUT NOT IN OVARIES.  
CC -I- DEVELOPMENTAL STAGE: LOW LEVELS IN 0-8 HOUR EMBRYOS AND ADULTS.  
CC HIGHER IN LATE EMBRYOGENESIS AND DURING LARVAL AND PUPAL STAGES.  
CC SHORT ISOFORM IS ENRICHED IN PUPAE AND ADULTS, LONG ISOFORM IN  
CC LARVAE.  
CC -I- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4  
CC subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.cn/announce>  
CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
CC -----  
CC EMBL; X89246; CAA61534.1; -;  
DR EMBL; Y15606; CAA75690.1; -;  
DR EMBL; AJ002073; CAA05172.1; -;  
DR EMBL; AE003667; AAF53914.1; -;  
DR EMBL; U36762; AAC46926.1; -;











QY	330	AYSEEDLSRALQDVYANKLDAKRSASCHNEHRSITLDRNLFRMKKHNDQSDHDSOELSEN	389
Db	255	DHLLEQLSDAVD--LISFL-----ORSCVD-----DEESTSPRODFE---	289
QY	390	DDAEAEVDSNASTR--VYPAEFAQAOLKTKLSHLESHNSDGLGEVDYR-GSPKMGHNPACG	446
Db	290	-----LVSTPRLTDSVTPVEQNHTNTTQDLVLENLLTQLTHNIVRKGSQQOQHONH	344
QY	447	NASANQAPASIRLDANVLHTLMLAAGIGAM-----PKLDETQVGDPIKGLLVANS-	499
Db	345	GVQQOQQOQHSHVQQOQQ--HNVOQOQHGVQQOQHVOQDERPYSQYHAT-----RGLMQOQRP	396
QY	500	--GGIMNEGGLNLLSASGENGNAGSALLQOQH--OQHNOQHQQOQHQQHVAAY-RH	553
Db	397	QHGGVQOQ-----AAIMSQOQOQLLSQOQOQHSHQOQOQHQAQOQHAAVQOQH	440
QY	554	RLPKSETPEPTNSSLDPNDASEDPIIKIPSFVSGSPASSSSLSPGVLGCHNPLNNNSL	613
Db	441	NIYVQQOQQOQHQQOQHQQH--HNNHQOQQOQRPQ	472
QY	614	SISNNSNHSNSHRNCSNRSFHSASPMILAAVQAGYSAGNSLLTSSSSSIQKMASSNIQ	673
Db	473	QSHSHSHHGHGHNSNMSLP--SPYIAAAAAAAAAAAAAAAAAAAAAA-----AAHLQ	520
QY	674	QIINEQSQEQLRGNGVYSDCSNNGSGSSSLGYKKPSIVAKIIGTDRSRGASBNLLSQ	733
Db	521	RPMSSSSSSGGTNSNNSGGSSN--SPLLDANAAARAALAL--DT--KPLIOV	569
QY	734	QHHSAHLHTHQOQOQOQLSAQALGKGTTPRKGRKRYNDRLSLVAVKAVORGEMS VBRAG	793
Db	570	SNPIGQPLNTQSQOQKQOQITLAKTTR-----YTF-----VENV-----	605
QY	794	SYGVGPHSTLEYKVKERHLMRKREKRPQDVLGLTGPANKLOLDIKAKGPHGGSKLNS	853
Db	606	-----SMDVTYKPELF-----SELKPEMTEIT--ABELTLEA-----	635
QY	854	ALKQNQNNQAAAAAAAAAAAAAPNGKILPLFAGQALSLFQRMFMPTQNMATN----	908
Db	636	-----ETTAATAAAAAAAAAAATTSATG-----GTQVLAAPRA--PLSGRKLGRGA	680
QY	909	---AYGLDFNRITEAMR-NPQA-----SNHHGLMKSADQMVENVYDGIIRKTLQASSEGGS	960
Db	681	KAVVYGGMTITLITLKSSEPVPTKTVHRTTSLATAAATAGLAPPTVSLNES	740
QY	1004	DYAATCSSASGES--VKRSGSPMGNYADIKERLSADSGGS-SPEEHSASHTNNNSLSLA	1066
Db	784	DCIDLDLHGHHSHQOQOQGSFQOQNMQOSVQOQSLQOQOQOQHSHNSNSAS	843
QY	1061	HNKNKSG-----GGGGCGGNGGTNGNCR	1083
Db	844	SNASHGSAEALCMGSSGGANEDSSGNNK	873



```

OY 467 HTLMLAGIGAMPKLDFTQVDFIKGLLVANSGIMNEGILNLILASQENSGNVAJLL 526
DB 525 KTSAVIA-----STTSSDR-----GGGLSHALTH--KVSPPSATAGRLV 565
OY 527 OOOOQHOOHQHOOOQOOQOAAVRRHL-----EKSTPEFNSSLDEND 571
DB 566 --EYTHOVHSPKRLILREFEKVSLDNGCVNNGSGAGGKRSPAKGTSTSSPAK 623
OY 572 ASIEDILIPFKVSGP---ASSSLSPGGLVGHHPPLNNNSLSI-----SNNSNMS 623
DB 624 AS--PMNLAPPGKRSPPSGSSSSSTSPATL--STOPRLNLSYISLSLGGSSGSS 678
OY 624 NSHRNGSRPSASPMLAAVAOGYSAKNSILTSSSSSIQKMAVNIQRIINBOGQE 683
DB 679 SSSSGGKKCGDH-----PAIISVHHHQHMYGPPSSSYRAL----- 717
OY 684 SLRNGNVSDCCSNNGS---SLIYKK--PSISAKII-----GCTDTSRFG 725
DB 718 -LTSKSPDVSGSGSGGKSPSHGTGKKRSPYSAGSPDYGSEFYRDPYAGRPSISG 776
OY 726 ASPNLISQHHSHHHLTHQOQOQOQSLAQELGKTRPKRGKRVNDRDSIYEAVALVGR 785
DB 777 SASODLSPRRS-----PASPATTPTVPKK-----TASIRRE 809
OY 786 EMSVHRAGSYGVPHSTLEYKVERHLMRPRKREKPPQPDVLGTGPANKQLDILKXGP 845
DB 810 FASPSASSSSCPSP-----GDRSASPPERHMQOQPHLQRRSPLHYVYPPPOYNG 861
OY 846 HGGSKLSNALKKQNNQAAAAAATAATP-----NGLKLPLEA 888
DB 862 NGSASPTSPAPTSNAAAAAATAATP-----NGLKLPLEA 921
OY 889 GPQALSPOPNMWPOTNTNAYGLDFNRTTEMRPQASNNHGLKMSAQDWENYDGI 948
DB 922 GASPL-LSPH---PPGSAALAA---AAAAARLSPOSAYH-----AFAYNGV 963
OY 949 RKTLOASGNGSAGNNGSNGNGH---GHGHGHGHALLDQLVKTPLPTPNRNDY 1005
DB 964 AAVAAAA---AAAFGCPAPSPHTPHLAPHQPH-----PALTTHTSPAH 1008
OY 1006 AAT-----CSSAGSEYKSGSPMGNYADIKREPLASDSCGSSDEE-HMSHINNNS 1057
DB 1009 LATPKLTDSTPDQMSATSSHRTASTSPSSSASASSASATSSAMFTHTSLNEOSS 1068
OY 1058 DLAHNKX 1065
DB 1069 DLEPLNSK 1076

```

```

CC specification of neuronal identity. May regulate transcription by
CC binding to DNA (By similarity).
CC -SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -SIMILARITY: BELONGS TO THE PROSPERO HOMEOBOX FAMILY.
CC -SIMILARITY: Contains 1 homeobox domain.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: AF190405; AAF06660.1; -.
CC FlyBase: FBgn028753; Dv1r\pros.
DR Pfam: PF05044; Prox1.
KW Nuclear protein; Transcription regulation; DNA-binding;
KW Homeobox; Developmental protein.
FT DOMAIN 4 12 POLY-ALA.
FT 19 36 ASN-RICH.
FT 150 172 ASN-RICH.
FT DOMAIN 206 209 POLY-ALA.
FT 237 292 GLN-RICH.
FT DOMAIN 305 309 POLY-ASN.
FT 349 381 SER-RICH.
FT DOMAIN 443 473 ASP-RICH.
FT 544 547 POLY-ALA.
FT 574 1080 GLN-RICH.
FT DOMAIN 888 906 ASN-RICH.
FT 1027 1030 POLY-ALA.
FT DOMAIN 1045 1054 POLY-GLN.
FT 1057 1062 POLY-GLN.
FT DOMAIN 1132 1189 THR-RICH.
FT 1140 1145 POLY-ALA.
FT 1154 1163 POLY-GLN.
FT DOMAIN 1183 1189 POLY-THR.
FT 1090 1097 NUCLEAR LOCALIZATION SIGNAL (BY
FT DOMAIN SIMILARITY).
FT 1330 1337 POLY-GLN.
FT 1394 1456 HOMEBOX (ATYPICAL) (BY SIMILARITY).
FT 1457 1556 PROSPERO-LIKE (BY SIMILARITY).
SQ SEQUENCE 1556 AA; 171029 MW; 6FACFEAD273B644 CRC64;

Query Match 4.1%; Score 249.5; DB 1; Length 1556;
Best Local Similarity 18.7%; Pred. No. 0.00011;
Matches 232; Conservative 167; Mismatches 434; Indels 405; Gaps 52;

OY 47 GSEDEPSQYNHSSKEISQSNFNHCKTENHRLQOQNGSQLLEEDSENNQTSRTPPT 106
DB 348 GSSNSSSHSNSSNNSTAS-----NSLRKSSDSDLSMQREERA---T 389
OY 107 PGATSTSPPEPPIDMRPSAKNFCVNGRLITVNAQG--KLVAESATATSS----- 157
DB 390 PATTVEPOLQNEVPPTKKEAVDMDLDEVELLGLSGHGRNSDLESILASPSHSDMLLDN 449
OY 158 -----TSNSHHQHDSDNSNSASLPHHSSSSSSNNNSG-----NNAR--HTAA 201
DB 450 SKDEVLDNDDIIEDEDDEDDDD--DCLLETREBNTSSNCLKPKGMELKRAYEHI-VS 506
OY 202 SARATPAATPANSLELYKLITTORAAKVTSMNSAAQLAOPSLADPULINSLSAQOQOQ 261
DB 507 SMRCSPSATTAQAGLOV-----NGCKKKRK-----YQPGQH 537
OY 262 OOOQIASAVPTTSEVSAALISPALKOTPPSPVDAPLDLSKSPSPNSISGDKSVARACA 321
DB 538 AMERYVAAALAGLNLQSMWLDQEBDESNELESF-OIOQKREKALKQLKSLSM----- 592
OY 322 TPTTSGRAYVEEPL-----GRALQDVANLKDAKSAQHHEQSIIDNRLFX 370
DB 593 -----OEQLAEMQOKYVOLCSHMEDESSCQELD-----NEHQDMDSHGE--Q 634
OY 371 MKHHDOEDHDGDELEDSNDDAEAVDSNASTPVYPAPAFARAKLKLH---LSBHNSGD 427

```

D	b		635	EDELQEDEGEPEEPLOQHNDIE-----LSHSPLTSKTPKE	671
O	y		428	LGEDVDRGSP-----KKGRHPACGNASANOAPASIPLDANVLLHTLM	470
D	b		672	SSEKAGSSSPYEATEEQOQLLRAAATMILPRPPICSLQWMSKMMSKLQHQNPVLGVAH	731
O	y		471	LAAGIGAMPKULDETQTGVEPFIKGLVANSGGIIMNEELNLYLASDGENSGNASILLQQOO	530
D	b		732	PALPGCFPELL---OHMGDMSHAAAAYQOFFEEOEARMAKEAEQOOO-----QQOO	780
O	y		531	HOOHHOOHQQOQQOOOONHAAYR-----HRLPK--SEPETNSSLDPNPA	572
D	b		761	QOQOQQOQQOQQOQQOQQOEQRRYEEDEEQRRMEQOQOLHRQOHLQOLLQOQQOQEHNVYA	840
O	y		573	SE-----DLPIKIFPSFKVSGPASSSSLPPGLVG-----HHHPILNNNSTLSI	615
D	b		841	AAVRPVQLPHRLPT-RWGAAAAAANAASHVLSKELSEKPKMLRQHNNSSSNSSNSSN	899
O	y		616	SNNSSHSSNSNR-NGSN-----RSPIASPMILAAY-----AOGYSAQSNTLL	657
D	b		900	NNNSNNNSNMWSMMSGTDLEGLADVLSKITTS--LSALDTITVFHFQRRLFISKADSV	957
O	y		658	TSSSSSIOK--WMASNI-----QRIINSGOGESLRNGVNVSOCSSSNGSSSL	703
D	b		958	TAAAQCLNLMDLWASQILDKRSPRTKVYAVDRGISGNGHPGAONGCPTPATQ-----SAAA	101:
O	y		704	GYYKPESIS-----VAKIIGTDTSRFGASP-----LISQGHSHAH	739
D	b		1013	MFAQEKTEPCGMNFVAAMALYNMSASFCLRPDQOQQOQQOQAQQOQQOQOVQCHQSHA-	1077
O	y		740	HLTHQOQQOQOLSAGEALKGTGPCKGKYNYRDSLVAEVKAVORGEMSVMRAGSYYGVF	799
D	b		1072	----QOTQOOLEQNREALSLVTTPK-----	109:
O	y		800	HSTLEYKVERMLMKRKREKRPQPDVLGTLPANKLOLDYLKAGPHGSKXLSMALKNON	859
D	b		1092	-----KKRKAKVTDIRITTPRVSRILLSGWRC-----AAOWTSCRSSCCRTTT	113:
O	y		860	NQAAAAAAAAAAAAAAAAATTENGCLKLFPE--AGPOLSFOPN---MFWPQTNATYAGL	912
D	b		1135	TTTTTAATAAATTATAATGCQOQOQQOQCNAGTENGERPGRRHLTLSTATTTTHDAGV	119:
O	y		913	DNFRITTEAMRNFOASNHGGLMKLSAADMVENVYDGIIKTLLQASEGNSAAGNSGNSGN	972
D	b		1195	-FAHLS-----GHTKSIAARIQSLF-AVOGITLO-----SACGAGR-----	122:
O	y		973	GHGHHGHGHALLDDLVVKKTPLPTPNRHNDYATATCSSASGESYKRSQSPMGNTADIAR	103:
D	b		1228	HGPRFHG-GOPTAACL-----HHQH-----HOMQLSSSPGSLCAL--	126:
O	y		1033	ERLSADSGSSSDEESHASHINNNSDSLANKKSGGGGGGGNGOT-----NGNGRSR	108:
D	b		1263	----MDSKSDPFLPHRPPTYHLHALLAHAH-----CGSPDYKTCILRAYMDADROSE	131:
O	y		1067	MTSRD---DSETDASSFKSGENGOQNHKMDLANGSS	1121
D	b		1311	CNSADMFTEDMQPTISFLK-----QOQOQQOQLSGAHS	1343
RESULT 10					
ID	YNF4_YEAST	STANDARD,	PRT,	1165 AA.	
AC	P53950;				
DT	01-OCT-1996	(Rel. 34, Created)			
DT	01-OCT-1996	(Rel. 34, Last sequence update)			
DT	01-OCT-1996	(Rel. 34, Last annotation update)			
DE	Hypothetical I28.1 kDa protein in OMP2-mg5 intergenic region.				
GN	YNL054W OR N2467 OR YNL2467W.				
OS	Saccharomyces cerevisiae (Baker's Yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
DC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_TaxID=4932;				

```

RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=5288C / FY1676;
RA MEDLINE=96021608; PubMed=8533472;
RA Berger P., Doignon F., Crouzet M.;
RT "the sequence of a 44 420 bp fragment located on the left arm of
RL chromosome XIV from Saccharomyces cerevisiae."
RL Yeast 11:967-974(1995).
[2]
RN ERRATUM.
RP MEDLINE=97060022; PubMed=8904343;
RA Berger P., Doignon F., Crouzet M.;
RL Yeast 12:297-297(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U12141; AAA9658.1; -
DR EMBL: 271330; CAA95925.1; -
DR PIR: S62982; S62982.
DR SGD: S0004999; VAC7.
DR GO: GO:0005737; C:cytoplasm; IDA.
DR GO: GO:0016021; C:integral to membrane; IDA.
DR GO: GO:0000324; C:vacuole (sensu Fungi); IDA.
DR GO: GO:0030234; P:enzyme regulator activity; IGI.
DR GO: GO:0006644; P:phospholipid metabolism; IGI.
DR GO: GO:0000011; P:vacuole inheritance; IMP.
SQ Hypothetical protein.
SQ SEQUENCE 1165 AA; 128140 MW; 0414978AA306CE7 CRC64;

Query Match 4.1%; Score 249; DB 1; Length 1165;
Best Local Similarity 19.8%; Pred. No. 8.1e-05;
Matches 210; Conservative 158; Mismatches 394; Indels 300; Gaps 50;

OY 107 PGATSTPPEPEPIEDWRPSAKCNCVNGRLTTVNAQKLVESAAT-----ATSSSTNS 161
DB 39 PSASTSTSPLEHREI-----VDSVAITAAITTSNVVQHNLPPTIDNNIMSDPAISHN 87
OY 162 HIHODSDSN-----SSASLP-----HHSSSSSSSNNNSGNGRARIHAASABATPA-- 208
DB 88 QDHH-SDINNAGTSMSTSDIPTDLHLEHIGSVSTNNNSNALLINPLSHLSNPSSS 146
OY 209 -----AATP-ANSLELYLTLTORAAKMTSMDSMAAQLAQSLLADFNILNLSIAS 256
DB 147 LRNNKSSLLVNASNPFAFSDVELSK--KKPAVISNNMPTSNIALYOTARSAIHGPSSSTA 204
OY 257 QQQQQQQQQQIASVATPTTSEVSAALISPAKLDTPSPVDAPLDDLSSKPSNHSISGVKS 316
DB 205 SKAFRRKSAFENNTAPSTSNNGTNTTP-----APL-----LPPLSLSQONKPKI 249
OY 317 VRACATPTPSGRARAYSEEDLSRALQDVANAKLDRKASQCHNEORSILDNLFPMKHHQ 376
DB 250 IERTPTAVYTNREILLENL---LDTRKAKAPAN---STHNDGPAVANDGLRIPNSNA 303
OY 377 EODHGDGELE-----DSNDAAEVDNSAST-----PVYPAEFARQRLKL--SH 419
DB 304 DDNENNNMMKNKKNKINSKGKERNDDTSCITTSKTPARSTAPLSTDNTOALTASVSSN 363
OY 420 LSEHNGSLDGEDVGRSGKMKGRHACGNASANOAPASIPLDANVVLHTLMLAIGIAMP 479
DB 364 ADNHNNN-----KKTSSNNNGNNSNSASNKTNDAIKUSMADLSSTSNNN 409
OY 480 KLD-----ETQVDEFIKGLIVANSQGINNEGILNLSAQEN-----SGNAGSL 525
DB 410 AINDSHESNEKRTKADFPAARLATAVG-----NEISDSEETVYESAGNSTKLI 462
OY 526 L---QQQHQHQHQHQHQHQHQHQVAAVYRHLRKSETPETNSSLDPNDASEDDILKIPS 582

```

```

Db 463 FPDSSSSQQQQQQQPPKQQQQQNN-----GITSKISA-----PLNNNK 502
Qy 563 FVSGPASSSSSLPGVLGGHHHPLNNNSLSISNNSSSHNGSRRSPHSAAPMLA 642
Db 503 KLLSLTKNSRHSITSTAIL-----NNTATISTNPNSNVWVNNNN----- 543
Qy 643 AAVAGGYSAGNSLITSSSSSIQKMASNIQRIQIENGQSESLRGNVSDCSSNGGSSS 702
Db 544 ---LMSGNHDEL-----SSIKQEPHQLOQO-----QPMVDQVSVDYSIDPDSNV 589
Qy 703 LGYKPS-----ISVAKIIGTDTSRFGASPVLISQOHSNAHL-----THQOQ---QQ 748
Db 590 IA-KSPDRKSSLSVLSKV-----SPHLSTSSNGNTISCENVATWSQLEPNN 637
Qy 749 QLSAGELGCKT--RPGKGKYRNY--DRDSLVEAYKAVQRGMSVHRAGSYGV-HSFL 803
Db 638 DISTKSLSNSTLRHSSANRNSNGDNKRLPTTYSKIFDSNPNGAPLPRYSGVDPHNL 697
Qy 804 EYKVERHLMRPKREKPPQPDVLGLTGPANKLQDKL-----KAGPHG---GSKLSNA 854
Db 698 EDYIQPH-----NYPTMQ-----NSVKKDEFPNSNNKFPHGLNFGD--NNV 739
Qy 855 LKQNNQAAAAAATAAATPENGKLPLEFAGPQALSFPQNNFWQTNATNAYGD- 913
Db 740 IEEENNCGSSNVNR-----FOHTNLQ--HEFIEDNESDENDIHS 777
Qy 914 ---FNRIEAMRNPOASNNHGLMKSAQDVENVYDGIIRKTLQASEGNSAAGNSNSN 970
Db 778 MFYNNKODLETYKPLISDY-----GEEDVDYD-----PNNATFNYSYSA 819
Qy 971 GNGHG---HGH---GHGHALLDQLLVKTPPLFTNRRNDYATCSASGESVKKSGSPM 1024
Db 820 SNTHELPLHGRNPSNSNDYDFWGNNG---NNQNLNEYPLAMKKGQRLSTNTNSI 876
Qy 1025 GNYADIKERLSADSGSSDEHSAHNNNSDLA---HN 1062
Db 877 MNG-----SIHNGNDVTHS---NINNNDIVGYSPPH 906

RESULT 11
BRC4_DROME STANDARD; PRT; 880 AA.
AC Q24206; Q46064; Q9W575;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Broad-complex core-protein isoform 6.
GN BR OR BR-C OR EG:17A9.1 OR EG:25D2.1 OR EG:123F11.1 OR
CG11493/CG11514.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A., DEVELOPMENTAL STAGE, AND CHARACTERIZATION OF
RP ISOFORMS.
RC TISSUE=imaginal disks, and larva;
RX MEDLINE=9629417; PubMed=8660872;
RA Bayer C.A., Holley B., Fristrom J.W.;
RT "A switch in broad-complex zinc-finger isoform expression is regulated
RT posttranscriptionally during the metamorphosis of Drosophila imaginal
RT discs";
RL Dev. Biol. 177:1-14(1996).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=Oregon-R;
RX MEDLINE=20196011; PubMed=10731137;
RA Benos P.V., Galt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA Dreno S., Gloux S., Leleau V., Mottier S., Galibert F., Bokkova D.,
RA Minana B., Kalatos F.C., Louis C., Siden-Klimos I., Bolshakov S.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo B.,

```

```

RA Model J., Peter A., Schoettler P., Werner M., Mourkioti F.,
RA Belnert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlancou A., Henderson N.S.,
RA McMillan P.J., Sales C., Tate E.A., Valenti P., Saunders R.D.C.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of D.
RT melanogaster";
RL Science 287:2220-2222(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dubin K.J., Evangelista C.C., Ferraz C., Ferrer S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Haxis N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liao P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Paclob J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.-N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [4]
RP CHARACTERIZATION OF ISOFORMS, AND MUTATIONAL ANALYSIS.
RX MEDLINE=97384928; PubMed=9242423;
RA Bayer C.A., von Kalm L., Fristrom J.W.;
RT "Relationships between protein isoforms and genetic functions
RT demonstrate functional redundancy at the Broad-Complex during
RT Drosophila metamorphosis";
RL Dev. Biol. 187:267-282(1997).
RN [5]
RP FUNCTION: BROAD-COMPLEX PROTEINS ARE REQUIRED FOR PUFFING AND
RP TRANSCRIPTION OF SALIVARY GLAND LATE GENES DURING METAMORPHOSIS.
RP SUBCELLULAR LOCATION: Nuclear.
RP ALTERNATIVE PRODUCTS:
RP Event:Alternative splicing; Named isoforms=6;
CC Name=6; Synonyms=BCORE-24;
CC IsoId=Q24206-1; Sequence=Displayed;
CC Name=1; Synonyms=BCORE-TN1-01-21;
CC IsoId=Q01295-1; Sequence=External;
CC Name=2; Synonyms=BCORE-01-21;
CC IsoId=Q01295-2; Sequence=External;
CC Name=3; Synonyms=BCORE-Q2-21;
CC IsoId=Q01295-3; Sequence=External;
CC Name=4; Synonyms=BCORE-22;
CC IsoId=Q01295-4; Sequence=External;

```





RA Jahnkecht R., Taube W., Luedcke H.-J., Pongs O.;  
RT "Characterization of a putative transcription factor gene expressed  
in the 20-OH-ecdysone inducible puff 74EF in *Drosophila*  
RT melanogaster.";  
RL Nucleic Acids Res. 17:4455-4464(1989).  
RN [2]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RX MEDLINE=90199900; PubMed=2107982;  
RA Butts K.C., Thummel C.S., Jones C.W., Karim F.D., Hogness D.S.;  
RT "The *Drosophila* 74EF early puff contains E74, a complex ecdysone-  
RT inducible gene that encodes two ets-related proteins";  
RL Cell 61:85-99(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,  
RA Abrial J.F., Agbayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Bernos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Botkova D., Botchan M.R., Boutck H., Brockstein P., Brotler P.,  
RA Butts K.C., Butam D.A., Butler J., Cadietu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Doseen C., Doup L.E., Dowes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hootin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,  
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Liao X., Lei Y., Levitsky A.A., Li Y., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclik J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Stimpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stepieton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*";  
RL Science 287:2185-2195(2000).  
RN [4]  
RP DEVELOPMENTAL STAGE.  
RX MEDLINE=94038699; PubMed=8223281;  
RA Huet F., Ruiz C., Richards G.;  
RT "Puffs and PCR: the in vivo dynamics of early gene expression during  
RT ecdysone responses in *Drosophila*.";  
RL Development 118:613-627(1993).  
RN [5]  
RL SUBCELLULAR LOCATION: Nuclear.  
CC -I- ALTERNATIVE PRODUCTS:  
CC Event=Synonyms=E74B;  
CC Name=B; Synonyms=E74B;  
CC Isoid=E11536-1; Sequence=displayed;  
CC Name=A; Synonyms=E74A;  
CC Isoid=E20105-1; Sequence=External;  
CC -I- DEVELOPMENTAL STAGE: In mid instar larvae salivary glands levels  
CC increase during 86-94 hours of development and represent the  
CC predominant isocorn during puff stage 1. Levels remain relatively  
CC constant in late larvae until the premetamorphic pulse of  
CC ecdysone. Transcripts are detected again from puff stages 12-14

```

CC      and 17-21.
CC      -I- INDUCTION: THE EXPRESSION OF THIS PROTEIN IS DEVELOPMENTALLY
CC      REGULATED AND IS CORRELATED WITH THE 20-OH-ECDYSONE INDUCED
CC      ACTIVITY OF PUFF 74EF.
CC      -I- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X15087; CAA33195.1; -.
DR      EMBL; M37083; AAA28494.1; -.
DR      EMBL; AE003523; AAF49324.1; -.
DR      PIR; S04722; S04722.
DR      HSRP; P14821; 25TT.
DR      TRANSFAC; T00210; -.
DR      FlyBase; FBgn0000567; Bp74EF.
DR      GO; GO:006914; P:autophagy; IMP.
DR      InterPro; IPR000418; Ets.
DR      InterPro; IPR002341; Hsf_ETS.
DR      Pfam; PF001079; Etsf; 1.
DR      PRINTS; PR00454; ETSDOMAIN.
DR      SMART; SM00413; ETS; 1.
DR      PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR      PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR      PROSITE; PS50061; ETS_DOMAIN_3; 1.
KW      Nuclear protein; Transcription regulation; DNA-binding;
KW      Alternative splicing; Developmental protein; Polymorphism.
FT      DOMAIN              78      280
FT      DOMAIN              281      759
FT      DOMAIN              760      883
FT      DNA_BIND            787      869
FT      DNA_BIND            27      51
FT      DOMAIN              72      79
FT      DOMAIN              289      299
FT      DOMAIN              372      378
FT      DOMAIN              406      415
FT      DOMAIN              434      437
FT      DOMAIN              486      489
FT      DOMAIN              510      524
FT      DOMAIN              614      630
FT      DOMAIN              639      642
FT      DOMAIN              651      661
FT      DOMAIN              695      703
FT      DOMAIN              707      711
FT      DOMAIN              712      717
FT      DOMAIN              721      733
FT      VARIANT             867      867
SQ      SEQUENCE            883 AA; 94820 MW; 148D5031A1BD1409 CRC64;
Query Match      4.1%; Score 244.5; DB 1; Length 883;
Best Local Similarity 20.4%; Pred. No. 9.6e-05;
Matches 177; Conservative 117; Mismatches 364; Indels 209; Gaps 28

145 LVASAAAT-ARSSSTNSHGHQHDSDSSASLPHHHSSSSSSNNNSGGR---ARHTAA 200
Db      7 LVAAAHNFASQAAALVNVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSATPPVAVSPVTP 66
QY      ||| : : : : :
201 AS---ARATPAAATPANSLELYKLTORAAKMTSNDMAAQLAOPSLADFNLIINSLSAQ 257
Db      67 TSPPPAAAAPAAASPPAAELQF-----DGGQAKTQEDPTMKDDQMLEKTQGE 114
QY      258 QQQQQQQQIQIASAVTPTTSVEASAAIISPAIKTTPSPVDAPLDLSKSP------NSSI 310
Db      115 VKDPVNVVEEPAIVDTESVMARQSPSPVASTKVEESLE---EISNKSPPVOEDEESESIV 171
QY      311 SGDYKSVKACATPTTPSGKRA--SEEDLSRLQDVVANKLDARKASQGHEDRSLTD--N 366
Db      172 ASDREKFFVLVNHLRQOQHHSFSSPDKTRSTLTDVSKILWERK---QQLQSSVITAAP 228

```





Query Match	4.0%	Score 243	DB 1	length 1507
Best Local Similarity	18.6%	Pred. No. 0.00021		
Matches 241	Conservative 157	Mismatches 433	Indels 462	Gaps 48

Qy	1	ERVAECCGRQWQMHYQDKLTCSHUNIEQO-----P1AIGSEDEPBOY	55
Db	461	EIIAQETKYEVTNPPIHISELOAKPQIOLSEKAEKTEIETKTATIPIVATATSTAO----	516
Qy	56	NHSSKEIISQSNPNH-----CKTENRLEQOHNSQLE-----EEDSENNQTSH	99
Db	517	---IKOLESPNPYKQIILQAEIILKREHNSPGGRITTAOLSSSSSGLAREEGRAPSVTA-	572
Qy	100	DSRTPTFGATSTPEP-----	115
Db	573	-SVLRSPAPRPLTPPTAVLCKKPTLGYEPNLPPTTAAIISSSNQOLQIAQQTOLQ	631
Qy	116	PEPIIDWRPAKCN-FCVNGRLT-----VNAGQ	143
Db	632	FOQPRQDNMSKFCSLFADDGKGLTYLKEEPDDLHSHLASTNCTQIDENTPPSDMVLGMLG	691
Qy	144	K-LVAESAATATSSSTNSHIOH-DSDSNSASLPHHIISSSSSSNNNSGCRARHIAAA	201
Db	692	TCLLPEDINSIDSTTCSFTASQHYQSPSSSSTAP---SWTSSSNNSYANSP--LSPL	745
Qy	202	SARAPRAATPRANSLELYKLTQRAAKTSMDSMAQLAQFSLDLPNLINSLSAQOQOQO	261
Db	746	TENSTATASNPBHOO-----QQHHNQOQOQO	771
Qy	262	QOQOQ-----IASAVPT--TSEVSAAAIAPALKDTPSPSPVADPLD--LSKSPS	305
Db	772	QOQOQHHPQHHDHNSSSNIDPLFYVYRESNNTSCSQHL-HSPSTSKSPBDSLSPLSCS	830
Qy	306	PNS-----SIGDVKSVRACATPTPSGRRAVEBVL---SRALQDVVANKLDAKASQO	356
Db	831	PNLSLOEBDDFSGTEAATMRAPYPIRIDDBPLLTETILMNCBPEDLQTNVKEIDALQOQLO	890

[illegible]

RA MEDLINE=92069760; PubMed=1720353;  
 RA Vaesens H., Geell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;  
 RA "Prospero is expressed in neuronal precursors and encodes a nuclear  
 RT protein that is involved in the control of axonal outgrowth in  
 RT Drosophila.";  
 RA Cell 67:941-953(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92171948; PubMed=1540176;  
 RA Matsumaki F., Koizumi K., Hama C., Yoshioka T., Nabeshima Y.;  
 RT "Cloning of the Drosophila prospero gene and its expression in  
 RT ganglion mother cells.";  
 RA Biochem. Biophys. Res. Commun. 182:1326-1332(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93083413; PubMed=1842358;  
 RA Chu-Lagrange O., Wright D.M., McNeill L.K., Doe C.Q.;  
 RT "The prospero gene encodes a divergent homeodomain protein that  
 RT controls neuronal identity in Drosophila.";  
 RA Development Suppl. 2:79-85(1991).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM C).  
 RX MEDLINE=20503846; PubMed=11051550;  
 RA Xu C., Kaufmann R.C., Zhang J., Klady S., Carthew R.W.;  
 RT "Overlapping activators and repressors delimit transcriptional  
 RT response to receptor tyrosine kinase signals in the Drosophila eye.";  
 RA Cell 103:87-97(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Berkeley;  
 RA MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaralides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blake R.G., Champé M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertam B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,  
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo J.M., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doull E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glöckle A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Labeo F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Relarte K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski R.P., Smith T.,  
 RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitzkas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RA Science 287:2185-2195(2000).  
 RN [6]  
 RP REVISIONS, AND ALTERNATIVE SPLICING.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=22426069; PubMed=12537572;

RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Betencourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter C., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review.";  
 RA Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [7]  
 RP SEQUENCE FROM N.A. (ISOFORM A).  
 RC STRAIN=Berkeley; TISSUE=Head;  
 RA Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.W.,  
 RA Champé M., Chavez C., Dorsett V., Farfan D., Frise E., George R.A.,  
 RA Gonzalez M., Guatin H., Li P.W., Liao G., Miranda A., Mungall C.J.,  
 RA Nunoo J., Paclab J.M., Paragas V., Park S., Phouanavong S.,  
 RA Wan K.H., Yu C., Lewis S.E., Rubin G.M., Celinker S.E.;  
 RA Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
 RN [8]  
 RP SIMILARITY TO C. ELEGANS CEH-26.  
 RX MEDLINE=94212446; PubMed=7909177;  
 RA Buerklin T.R.;  
 RT "A Caenorhabditis elegans prospero homologue defines a novel domain.";  
 RA Trends Biochem. Sci. 19:70-71(1994).  
 RT -!- FUNCTION: Required for proper neuronal differentiation of most or  
 CC all neurons and their precursors in central and peripheral nervous  
 CC systems, axonal outgrowth and pathfinding. Not required for the  
 CC specification of neuronal identity. May regulate transcription by  
 CC binding to DNA.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Comment=Experimental confirmation may be lacking for some  
 CC isoforms;  
 CC Name=C; Synonyms=L;  
 CC IsoId=P29617-1; Sequence=Displayed;  
 CC Name=A;  
 CC IsoId=P29617-2; Sequence=VSP\_002306, VSP\_002307, VSP\_002308;  
 CC Name=B;  
 CC IsoId=P29617-3; Sequence=VSP\_002307, VSP\_002309;  
 CC Name=S;  
 CC IsoId=P29617-4; Sequence=VSP\_002308;  
 CC -!- TISSUE SPECIFICITY: NEURONAL PRECURSORS, EXPRESSED IN THE  
 CC DEVELOPING CNS. LENS-SECRETING CONE CELLS OF THE EYE, AND MIDGUT.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN NEURONAL PRECURSORS EARLY DURING  
 CC FORMATION.  
 CC -!- SIMILARITY: BELONGS TO THE PROSpero HOMEOBOX FAMILY.  
 CC -!- CAUTION: Ref.7 sequence differs from that shown due to a  
 CC frameshift in position 1122.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M81389; AAA28841.1; -;  
 CC EMBL; D10609; BA01464.1; -;  
 CC EMBL; Z11743; CA77802.1; -;  
 CC EMBL; AF190403; AAF05703.1; -;  
 CC EMBL; AE003691; AAF54628.2; -;  
 CC EMBL; AE003691; AAN13500.1; -;  
 CC EMBL; AE003691; AAN13501.1; -;  
 CC EMBL; AY060680; AAL28228.1; ALT\_FRAME.  
 CC PIR; S24548; S24548.  
 CC PD3; IMIJ; 04-DEC-02.  
 CC FlyBase; FBgn004595; pros.  
 CC GO; GO:0045179; C:apical cortex, IDA.  
 CC GO; GO:0045180; C:basal cortex, IDA.  
 CC GO; GO:0005634; C:nucleus, IEP.

[illegible]



Db 1755 -YSLNGKLSANGVGGSSGSSGSSAAGNGSGGSSNGVLGGSSGSSNSGNN 1813  
QY 737 SAHILTHQOQ---QQQLSAQELAKG-----TPKRGKRYNYRDSL 776  
Db 1814 SSHHKTEDQQNDHDLHAGSKLLGSSNFMIAKHPVGLAIKEITRNKPSSESSQMTSDT 1873  
QY 777 EAVKAV--QRGEMSV-----HRAGSYG--VPHSTLEKVKER--HLMRPKREPK--POP 824  
Db 1874 ESCETLHNRNOMAVLAMEHTAKELNGSHAHGCHHQQPQRTTHORPRSKELQYSHE 1933  
QY 825 DLVGLTGPANKLQLDKLYAGPHGSKLSNALKNQNNQAAAAAAAAAATPGLKLP 884  
Db 1934 SLDDG---AVQSQSQ----- 1948  
QY 885 LFEAGPOLSPQPMFWPQTNTATNAYGLDFNRITAMRNPOASNHGMLKSAQDMVENY 944  
Db 1949 -----RNQRYHHHHHQROOQOQRYNHVQOEERDPTEDNLA 1986  
QY 945 D-----GIIRKTLQASEGNSAAGNSGNSGNGHGHGHALDQLLVKTP 994  
Db 1987 DEEFDEDEYGRDVRQKRLQKSE----- 2008  
QY 995 LPFTNRRNDYATCSASGSEVKRSGSPMGNYADIKERLSADSGSDEEHSASHINN 1054  
Db 2009 ---LNHKRSEVA-----TEAGNHDEVE--EEDDDDEEDHR----- 2042  
QY 1055 NNSDLAHNKNKSGGGGGGQNGNGRSGSRMTSRDSEDTASSFKSGENGGOONHKM 1114  
Db 2043 -----NGGREAAPLTNGSMRGLNANVIND-----ELKYGAT--HLMHQM 2080  
QY 1115 DLNGSSSS--SHIKSEEAATGHS-----PGHHTTSLHE 1149  
Db 2081 DSNPLESOSSEWSDDCREBATGAESTGYTTDEPLENISLINE 2124

Search completed: October 29, 2003, 12:28:18  
Job time : 40 secs



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 107007

TO: Minh-Tam Davis  
Location: cm1/8a01/8e12  
Art Unit : 1642  
Wednesday, October 29, 2003

Case Serial Number: 10/016768

From : Susan Hanley  
Location: Biotech-Chem Library  
CM1 6B05  
Phone: 305-4053

susan.hanley@uspto.gov

### Search Notes

**RUSH**

*pty date:*  
*10/2000*



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 12:25:06 ; Search time 126 Seconds  
(without alignments)  
2385.963 Million cell updates/sec

Title: US-10-016-768a-10

Perfect score: 6030

Sequence: 1 MHSSYSLESLERVAEECMGR.....ILHEKLAQIKAEQVDADQL 1165

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rident:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6014	99.7	1165	5 Q9VD60	Q9VD60 drosophila
2	4839	80.2	1221	5 Q24079	Q24079 drosophila
3	707.5	11.7	1598	5 Q95YM8	Q95YM8 apis mellif
4	336	5.6	2472	5 Q8T2M5	Q8T2M5 dictyosteli
5	329.5	5.5	1918	5 Q8SSW3	Q8SSW3 dictyosteli
6	314.5	5.2	2151	5 Q9VPL6	Q9VPL6 drosophila
7	314.5	5.2	5322	5 Q9VPL9	Q9VPL9 drosophila
8	311.5	5.2	2112	5 Q9VEL9	Q9VEL9 drosophila
9	310.5	5.1	2038	5 Q9W1L3	Q9W1L3 drosophila
10	309.5	5.1	2151	5 Q9Y1L3	Q9Y1L3 drosophila
11	308.5	5.1	1103	5 Q9VY72	Q9VY72 drosophila
12	305	5.1	1444	5 Q9VTN2	Q9VTN2 drosophila
13	305	5.1	1514	5 Q8S1S5	Q8S1S5 drosophila
14	304	5.0	1249	5 Q8INC6	Q8INC6 drosophila
15	302	5.0	1140	5 Q8T1T0	Q8T1T0 dictyosteli
16	302	5.0	1565	5 Q8T1M2	Q8T1M2 dictyosteli

17	296	4.9	1162	5 Q9W3Q5	Q9W3Q5 drosophila
18	295.5	4.9	1508	5 Q24341	Q24341 drosophila
19	293	4.9	1456	5 Q9VX23	Q9VX23 drosophila
20	293	4.9	3381	2 Q9KX33	Q9KX33 streptococc
21	293	4.9	4001	5 Q8WRQ7	Q8WRQ7 drosophila
22	292	4.8	2310	5 Q9GR49	Q9GR49 drosophila
23	290	4.8	1880	5 Q8MP27	Q8MP27 dictyosteli
24	287	4.8	1404	5 Q8IRY3	Q8IRY3 drosophila
25	285.5	4.7	939	5 Q9NH00	Q9NH00 drosophila
26	285.5	4.7	1409	5 Q9VX60	Q9VX60 drosophila
27	285	4.7	1721	5 Q8SSQ0	Q8SSQ0 dictyosteli
28	283.5	4.7	1180	5 Q24163	Q24163 drosophila
29	283	4.7	1416	5 Q9VYM1	Q9VYM1 drosophila
30	282.5	4.7	1741	5 Q9W517	Q9W517 drosophila
31	282.5	4.7	1768	5 Q24153	Q24153 drosophila
32	281	4.7	929	5 Q9HBM0	Q9HBM0 drosophila
33	281	4.7	2280	5 Q9VBE6	Q9VBE6 drosophila
34	281	4.7	2302	5 Q9N693	Q9N693 drosophila
35	280.5	4.7	1741	5 Q46095	Q46095 drosophila
36	280	4.6	929	5 Q9NBL3	Q9NBL3 drosophila
37	279.5	4.6	929	5 Q9NGW5	Q9NGW5 drosophila
38	279.5	4.6	1186	5 Q9VED3	Q9VED3 drosophila
39	279	4.6	2150	5 Q23863	Q23863 dictyosteli
40	279	4.6	2181	5 Q9VRA6	Q9VRA6 drosophila
41	278.5	4.6	1673	5 Q9VQ19	Q9VQ19 drosophila
42	278.5	4.6	1785	5 Q81520	Q81520 plasmodium
43	277.5	4.6	1329	5 Q9W283	Q9W283 drosophila
44	277.5	4.6	2283	2 Q8VQ99	Q8VQ99 staphylococ
45	274	4.5	1338	5 Q8KN23	Q8KN23 dictyosteli

#### ALIGNMENTS

RESULT 1  
Q9VD60 PRELIMINARY; PRT; 1165 AA.  
AC Q9VD60;  
DT 01-MAY-2000 (TRENBLREL. 13, Created)  
DT 01-OCT-2002 (TRENBLREL. 22, Last sequence update)  
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)  
DE CG18389 protein.  
GN EIP93F OR CG18389.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brannon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Bortman M.R., Bouck J., Brockstein P., Broctier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadelu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
RA Foster A., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Gloeck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harkin N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,





[illegible]

```

Db      1181 VPTLPEYFATSKIRGLQEQGRRAAWVQQQQQQQQQQQQQQQQQQQQQQQQQQQ-----QQQHQA 1236
Qy      955 SEGNGSAAG-----NGSNGSNG-----NGHGHHGHALLDOLVK--KTPL 995
Db      1237 RERGVGAGIAETSGATSNRSGAQMSKVPRDVGEGIYDGGANGSFLDNLIRSLGTGI 1296
Qy      996 P-----FTNNHNNNYAATCSSASESVKRSQSPMGNVADIRE----RLSDSGGSPDE 1046
Db      1297 PRDQPMATEANQQQQQASSQQQIPESM-RSVALIDLCRNSRRTPVPLADS--SEDES 1353
Qy      1047 HSASHINNNS-----DLAHNKNKSGGGGGGGGCGQTGNCRSSRMTS----- 1089
Db      1354 YRGSSASGCRVPRPERPVTVLDSPESSDGRNDG-----SDRLTSPPTPLS 1402
Qy      1090 -----RDSETDASS-FKSGENGQQQNHK 1112
Db      1403 ISRAGSRDESDTSDTKLDRSSREREVNHGQQQEDR 1438

RESULT 4
Q8RTM5      PRELIMINARY;      PRT; 2472 AA.
ID      Q8RTM5
AC      Q8RTM5;
DT      01-JUN-2002 (TremBtrel. 21, Created)
DT      01-JUN-2002 (TremBtrel. 21, Last sequence update)
DT      01-OCT-2002 (TremBtrel. 22, Last annotation update)
DE      Hypothetical 278.5 kDa protein.
OS      Dictyostelium discoideum (Slime mold).
OC      Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX      NCBI_TaxID=44689;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=J4;
RA      Gloeckner G., Elchinger L., Szafranski K., Pachebat J., Dear P.,
RA      Lehmann R., Baugart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA      Tungal B., Cox E., Quail M.A., Platzer W., Rosenthal A., Noegel A.A.;
RT      "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL      Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
DR      EMBL; AC115593; AAL92295.1; -.
KW      Hypothetical protein.
SQ      SEQUENCE 2472 AA; 278497 MW; 30CCFF7157D4008A7 CRC64;

Query Match      5.6%; Score 336; DB 5; Length 2472;
Best Local Similarity 18.3%; Pred. No. 2.8e-10;
Matches 206; Conservative 172; Mismatches 33; Indels 356; Gaps 35;

Qy      38 EEOQPIATAGSEDEPSQYNHSSKEISQSPYHCKTEHNRLEQHNQSOLLEEDSENNQT 97
Db      645 QERDDTVYLGSLIEFDVNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 704
Qy      98 SHDSRRTPPGATSTPSPPEPIDWRPQKCNFCVNGRLTLVNAQGKLVAESAATATSSS 157
Db      705 NNNNNNN-----NNNNNN-----NNSNS 717
Qy      158 TSNHHIOHDDSNSSASLPHHISSSSSNNSSNGNRARHIAAASARATPLAATPANSLE 217
Db      718 SSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 775
Qy      218 LYKLLTQGAAMTSMDSMAAQLAQFSLADENLNSLASQQQQQ-----QQQQAISAVTPT 273
Db      776 -----NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 806
Qy      274 TSEVSAALISPALKDTPSPVDADLDSLSSKSPSSISGDVKSRYACATPTPSGRRAYSE 333
Db      807 ODE-----SITRELSSRRNSSGGLINSSRRNNNNNNNNNNNNNNNNNNNNNNNNNNNN 857
Qy      334 EDLSRALODVANKLDAARKSASQHHQEQRSILDNRLFKWKHHHDQFODHDGDELESDNDAE 393
Db      858 -----LQNI-----NNSYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 898
Qy      394 AEVDSNASTPYVPAEFAAQJRKLSHLSHNHSGDGLGEVDGSGKMGNRACGNAASNG 453
Db      899 ---INNSQVPLIP-----NTVNNINPNADPPRYVEVD--QYKTLIGHPLTRVSSPEST 946

```







Db 1463 OQOAGNEAAGAAASASTGSGAGGKAAASASOLAAGFPFLPNSLTPPWGLGGL 1522  
 Qy 761 RP----KAGYRNYDRDLEAVEAVKAVORGEMSVHAGSYGVPHSTLEKVKERHLMRP 816  
 Db 1523 NPYSLGSGGSLGSAVDQ--LAQOYNILNGATTSASTSTQSKSHOSKSSOSRRTTASA 1580  
 Qy 817 KREPKPPDVLGTLPANKLQDLKLAGPHGSGKLSNMLKNQNNQAAAAAASAAAAA 876  
 Db 1581 NSASLMMAMAMSGCASTVTTPTTSAGSGRGGRSS--RNG--SOTTTTAAADQLSSLL 1638  
 Qy 817 TPNGIKLFLFEGAP---QALSPQPMWFPQ--TNATNAYGL-----DNRITAMR 922  
 Db 1639 MPG-----ADPHLESLSRMSNMDLAQATRLMSSLGMPPLSGTSPSSCGGNTSKR 1690  
 Qy 923 NPOASNNHGLM--KSNQDVENYVGGIIRKTLQASGNGNSAAGNGSNGNGHGHGHG 981  
 Db 1691 SSQAANENAVQAKGQKWLKSLARCALPTDLAALQAFSGCKMPSTSGSN----- 1739  
 Qy 982 HALLDQLLVKTPPLPFTNHRNDYATCSASGESVKSQS--PMGNADIKERLSADS 1039  
 Db 1740 -----TGTSSTSSKSKAAATAAAQLPPIPMSSDPQAFLEMAAQAAMA--A 1787  
 Qy 1040 GGSDESHASHINNNSDLAHNNKXSGCGGCGGNGGTNGGRSSRMTSRDSETDASS 1099  
 Db 1788 GGSILPLSGPGS-----LASLAGLTGSGAGGCG--SSASGTSSTSSSKRQREDA-- 1834  
 Qy 1100 FKSGNGGQONH--KMDLN-----CGSSSSSHIKCESAALGCHSPGHHTTSI 1146  
 Db 1835 FKQ-----QMDYVYKTLGLSGGSLIPTSSAGSGSSAANAAAAA-- 1878  
 Qy 1147 LAEKLAQIKAEQVDQADQ 1164  
 Db 1879 -----AALDAEQHQOQ 1891  
 RESULT 7  
 Q9VPL9 PRELIMINARY: PRT: 5322 AA.  
 AC Q9VPL9; 09N164;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE CG3696 protein (KISMET-L long isoform).  
 GN KIS OR CG3660 OR CG3696 OR CG38326.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blasej R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt R.G., Nelson C.R., Miklos G.L.G.,  
 RA Abri'l J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,  
 RA Butts K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA LaSato P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mervin G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celinker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Buesam D.A.,  
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Fartan D.,  
 RA Ferreria S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.B., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutnák P., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Therrien M., Morrison D.K., Wong A.M., Rubin G.M.;  
 RT "A Genetic Screen for Modifiers of a KSR-Dependent Rough Eye Phenotype  
 RT in Drosophila.";  
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AE003590; AAFF1527.3;  
 DR EMBL; AF215703; AAF33004.1;  
 DR HSSP; P23197; IAP0.  
 DR FlyBase; FBGN001309; kis.  
 DR InterPro; IPR000953; Chromo.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase.  
 DR InterPro; IPR000330; SNF2\_N.  
 DR Pfam; PF00271; helicase\_C\_1.  
 DR Pfam; PF00176; SNF2\_N\_1.  
 DR SMART; SM00298; CHROMO; 2.  
 DR SMART; SM00487; DEXDC; 1.

OY	96	OTSHDSRPTPGASTPSTPRPEPIDMRPSACNFCVNRILLITVAOGLVESAATTS	155
Db	3847	OTSHIQOYASSASASTSGVDPVOTQSKPS-----SLSRLGLAKSTINAAVAABA	3900
OY	156	SSTSNSHITHOHSDSNSASLPHHISSSSSNNNSG-----NPARHIA-	199
Db	3901	AAGN-----ATSLSSLLPGMSLGAATSSAGVSGLSVPBAGSVGKKRHHIAI	3955
OY	200	-AASARATPAAATPANSELYKLTGORAAMTSMDSMAQLAOFSLADFNL--INSLAS	256
Db	3952	DVETERA-----KLHALLNSMTAPKDMSESIANMEALGSGGGRGNSASS	4000
OY	257	OOQOOOOOOLASVPTTSEVAAALSPALKTPEPVDAPLIDLSK-----PSP	306
Db	4001	GMQPPRAHOHSLSSQSGOFKSPAV-PAMK-TPPSSMGAPWIDLSSTSPKMMTEMLXA	4056
OY	307	NSISGDVKSVCATPT-----PS-GRRAYSEEDLSRALODVANKLDARK	352
Db	4059	SSSGALDLSVQDFPMFPGKXSVMHAALSAFFPSMGNKSLDITLKLKMK--KNCTIE	4111
OY	353	SASQNHQEPSLIDNLFPKYKHHDOBDHDELEDSNDDEAEVDSNMS-TPVPAEFAR	411
Db	4116	PVIGKEKRRKXKLDIVLGLSAKKQKTFEPDPLPSKK--KQIIPPSVTP-----	4166
OY	412	AQLRLKLSLSEHN-----GSDLDGEDVDRGSPKXGRHAPAGNSANGARA----	456
Db	4165	ANLQSSSSNOQSKQKFTTIVTTVPKSKSGSSNSGSGTGSSASAGGSGGLSALOM	4222
OY	457	-----SIPLDANVL-----HTLMLAIGAMPKX-481	
Db	4225	AMGSLSKXDSLNALLAQMTADPQTFKQOQGMQFLPPAQKAYENMLAEHQAM-KIS	4283
OY	482	-----DETQVQDFIKGLLVANSQGITMEGLINLLSASQENSNGMS	523
Db	4284	SKFSTNSPHDVKNKMLSDMTSPILDQOL-SIDYVGSGGSGSGSGNSRKSNNQOQNSQS	4344
OY	524	LLLOQOQHQQHNOHQOQOQOQNYAAVNRKPSBT-----ETNSSLDPNDA-----	572
Db	4343	SSAAOLOQXQOQOQOQOQOQOQSHSPGPNLTGEBEVPVITNKQTKRKLQGNAPOLKRLM	4400
OY	573	---SEDPILKI-PSF--KVSQPASSSLSPGGLVGCH-----HHRNNNSLISIN	617
Db	4403	QWLTENPVEVDPKMLEQOMQNMSPSPRPRASMESGYGSAKXSHQGRPLSNLSTSSSS	4466
OY	618	NSNHSNSGHRN--GSN-----RSPHSASPMILAAVAAGGYSAQ-----	653
Db	4463	HTQOQSSAAQSGAGNSGSSKKNSRQQTAAALDQALLOFSLGLNPSLIANTLPGIGA	4522
OY	654	-----NSLITSSSSSIQKMAASNIQOQINBSQGESLRNGVSPCCSSNNGGS	700
Db	4523	FDPKRPPLAFDPRKPLSWSFGGMPRM-GNI-----PGJNINNNMLFASLGMGGL	4573
OY	701	SSL-GYKKPSI-----SVAKITIGFDITSRFGA-----SPNLSQOHNASHNLTHOOQ	746
Db	4574	GNLACMDPOSLLAALMAAGPTLGLTGASGAGAGSGKSOAQOSQSSATSSSSSKKKKQOOQ	4633
OY	747	QOQLSAQEAL-----GK-----GT	760
Db	4634	QOQAAQNEAAQALAAASASTGSGAGGKNAALASQALAGPFLFPNPFLPYRPMGLGL	4693
OY	761	RP-----KXGKYNNYPRDLSVEAVKAVQREMSVNHAGSYUYGRNHTLEKKUVERHLMRR	816
Db	4694	NPYSIGSSGLGSAVQO--LDAQOYNLLNGATSSASNTSTQOSKHOSQOSKQSSRNITGASA	4755

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
PA Mount S.M., Murthy P., Murthy I., Muzny D.M., Nelson D.T.,

RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Paclel J.M.,

Query Match	5.2%	Score 311.5;	DB 5;	Length 2112;
Best Local Similarity	18.9%;	Pred. No. 5.6e-09;		
Matches 225;	Conservative 167;	Mismatches 411;	Indels 389;	Gaps 41

[illegible]

Oy  
1099 SFFSGENGCCQNNHMKMDLNGSSSS-----SHIKCESAPATGHHSPGHHTTS 1145

Ddb  
1069 SSSSSSNGCGSSNTGTSSNSNGASSGGGSNOCSSSNSGSSGGSSNGCQNSTS 1120

RESULT 9	09WJL3	PRELIMINARY;	PRT;	2038 AA.
ID	09WJL3			
AC	09WJL3;			
DT	01-MAY-2000 (TREMBlrel. 13, Created)			
DT	01-OCT-2002 (TREMBlrel. 22, Last sequence update)			
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)			
DE	CG2252 protein.			
GN	FS(1)H OR CG2252.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxId=7227;			
XP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Berkley.			
RC	MEDLINE=20196005; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abdill J.F., Agbayan A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,			
RA	Ballew R.M., Baau A.V., Bakendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beehon K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Cencar A., Chandra I.,			
RA	Cherry J.M., Casaley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hoschkin D., Houston K.A., Howland T.J., Wei M.-H., Idegwu C.,			
RA	Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Laeko P., LeY., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Milhina N.V., Mobarry B., Murphy L., Muris D.M., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muris D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Patel J.M.,  
 RA Palazolo M., Pictman G.S., Pan S., Pollard J., Puri V., Reen M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Stapleton A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtsek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Barton J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Chape M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farrant D.,  
 RA Ferrelia S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibbegam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Patel J., Paragis V., Park S., Patel S., Pfeiffer B., Nunoo J.,  
 RA Phouanavong S., Pictman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svrtsek R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 [3]  
 RN SEQUENCE FROM N.A.  
 RP Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hirdedy P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Bernat B., Carlson J.W., Ceiniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn G., Richter J., Ruso S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 [5]  
 RN SEQUENCE FROM N.A.  
 RP FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL, AE003442; AAF46312.3; --  
 DR HSSP; Q92831; 1891.  
 DR FlyBase; FBgn004656; fs(1)h.  
 DR InterPro; IPR001045; Antifreeze 1.  
 DR InterPro; IPR001487; Bromodomain.  
 DR InterPro; IPR002173; PFKB.  
 DR Pfam; PF00439; bromodomain; 2.  
 DR PRINTS; PR00308; ANTIFREEZE1.  
 DR PRINTS; PRO0503; BROMODOMAIN.  
 DR SMART; SMO0297; BROMO; 2.  
 DR PROSITE; PS00633; BROMODOMAIN 1; 2.  
 DR PROSITE; PS00633; BROMODOMAIN 2; 2.  
 DR PROSITE; PS00583; PFKB\_KINASES\_1; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_2; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_3; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_4; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_5; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_6; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_7; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_8; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_9; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_10; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_11; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_12; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_13; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_14; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_15; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_16; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_17; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_18; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_19; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_20; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_21; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_22; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_23; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_24; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_25; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_26; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_27; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_28; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_29; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_30; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_31; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_32; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_33; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_34; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_35; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_36; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_37; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_38; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_39; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_40; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_41; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_42; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_43; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_44; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_45; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_46; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_47; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_48; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_49; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_50; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_51; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_52; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_53; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_54; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_55; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_56; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_57; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_58; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_59; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_60; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_61; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_62; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_63; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_64; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_65; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_66; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_67; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_68; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_69; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_70; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_71; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_72; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_73; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_74; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_75; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_76; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_77; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_78; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_79; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_80; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_81; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_82; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_83; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_84; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_85; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_86; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_87; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_88; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_89; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_90; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_91; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_92; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_93; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_94; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_95; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_96; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_97; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_98; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_99; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_100; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_101; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_102; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_103; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_104; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_105; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_106; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_107; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_108; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_109; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_110; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_111; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_112; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_113; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_114; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_115; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_116; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_117; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_118; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_119; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_120; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_121; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_122; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_123; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_124; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_125; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_126; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_127; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_128; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_129; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_130; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_131; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_132; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_133; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_134; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_135; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_136; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_137; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_138; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_139; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_140; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_141; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_142; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_143; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_144; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_145; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_146; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_147; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_148; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_149; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_150; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_151; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_152; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_153; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_154; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_155; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_156; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_157; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_158; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_159; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_160; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_161; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_162; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_163; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_164; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_165; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_166; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_167; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_168; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_169; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_170; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_171; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_172; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_173; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_174; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_175; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_176; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_177; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_178; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_179; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_180; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_181; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_182; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_183; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_184; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_185; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_186; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_187; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_188; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_189; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_190; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_191; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_192; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_193; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_194; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_195; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_196; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_197; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_198; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_199; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_200; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_201; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_202; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_203; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_204; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_205; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_206; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_207; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_208; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_209; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_210; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_211; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_212; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_213; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_214; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_215; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_216; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_217; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_218; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_219; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_220; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_221; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_222; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_223; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_224; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_225; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_226; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_227; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_228; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_229; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_230; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_231; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_232; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_233; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_234; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_235; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_236; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_237; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_238; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_239; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_240; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_241; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_242; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_243; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_244; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_245; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_246; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_247; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_248; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_249; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_250; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_251; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_252; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_253; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_254; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_255; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_256; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_257; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_258; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_259; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_260; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_261; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_262; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_263; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_264; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_265; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_266; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_267; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_268; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_269; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_270; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_271; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_272; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_273; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_274; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_275; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_276; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_277; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_278; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_279; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_280; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_281; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_282; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_283; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_284; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_285; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_286; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_287; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_288; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_289; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_290; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_291; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_292; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_293; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_294; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_295; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_296; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_297; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_298; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_299; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_300; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_301; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_302; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_303; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_304; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_305; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_306; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_307; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_308; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_309; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_310; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_311; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_312; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_313; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_314; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_315; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_316; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_317; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_318; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_319; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_320; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_321; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_322; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_323; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_324; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_325; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_326; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_327; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_328; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_329; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_330; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_331; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_332; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_333; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_334; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_335; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_336; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_337; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_338; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_339; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_340; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_341; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_342; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_343; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_344; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_345; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_346; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_347; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_348; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_349; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_350; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_351; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_352; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_353; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_354; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_355; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_356; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_357; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_358; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES\_359; 1.  
 DR PROSITE; PS



RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Aamatiades P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zheng Q., Chen L.X.,  
 RA Branton R.C., Rogers Y.H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt C., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernick B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Bortman M.R., Bouck J., Brockett P., Broctier P.,  
 RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dubin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,  
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Metkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sideris-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier B., Spirdling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Weissman D.A., Weinstein G.M., Weissenhach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 [12]  
 RN SEQUENCE FROM N.A.  
 RP Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Aamatiades P.G., Branton R.C., Rogers Y.,  
 RA Batzon J., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwan C., Jatali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
 RA "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 [13]  
 RN SEQUENCE FROM N.A.  
 RP Mirra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminke J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Bernan D., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 [14]  
 RN SEQUENCE FROM N.A.  
 RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 [15]  
 RN SEQUENCE FROM N.A.  
 RP FlyBase;

RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: AE003494; AACF48333.2; -!  
 DR FlyBase: FBgn0052611; CG32611.  
 DR InterPro: IPR000104; Antifreeze\_1.  
 DR InterPro: IPR000210; BTB\_POZ.  
 DR InterPro: IPR007087; ZnF\_C2H2.  
 DR InterPro: IPR007086; ZnF\_C2H2\_sub.  
 DR Pfam: PF00651; BTB; 1.  
 DR Pfam: PF00096; ZF-C2H2; 4.  
 DR PRINTS: PR00308; ANTIFREEZE1.  
 DR PRINTS: PR00048; ZINC\_FINGER.  
 DR SMART: SM00225; BTB; 1.  
 DR SMART: SM00355; ZnF\_C2H2; 4.  
 DR PROSITE: PS00097; BTB; 1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
 DR PROSITE: PS01577; ZINC\_FINGER\_C2H2\_2; 4.  
 DR Metal-binding; Nuclear protein; zinc; zinc-finger.  
 SQ SEQUENCE 1103 AA; 113716 MW; EB586FD5683C9B0B CRC64;  
 Query Match 5.1%; Score 308.5; DB 5; Length 1103;  
 Best local similarity 20.3%; Pred. No. 3.6e-097;  
 Matches 221; Conservative 130; Mismatches 366; Indels 373; Gaps 40;  
 147 AESAATATSTSTNSHIHQHDSNSASLPFHITSSSSSSNNNSGNNRARIIAASARAT 206  
 120 ATQATATATATATAA-----ATSTSTATSAATTAATATATATATATATATAT 160  
 207 PAATPANSLELYLTLORAAKMTSMDSMAAQLAFSLADFNLIINSIASQQQQQQQQQ 266  
 161 ATAAAGSSN-----TTTTTAATTAATCATAA-----T 187  
 267 ASATPTTSEVSAAAIAPALDTPSPVDAPLDLSKSPNSISGDKVYACATPTPS 326  
 188 STAATSSSSSTSAATAA---AAASAGVAHEEATSSSTSGCGKREASDSSSTPPA 243  
 327 GRAYSEEDLSRALQDVVANKLARKASQCHNEORSILDNRLFKMKHHDEQDHDCLE 386  
 244 KRSSRMHP-----ADKVDAEDREQONQADELSSELLGRNLKBEEDDVDEL 293  
 387 DS-----NDDEAEVDSNASTPVY-----PAEPRAQLRKLSHS----- 421  
 294 ENERTKLGRDEDDDEDEEDTPEPLDLYORPNVEPKSAATAAAGGSSNSISGSNA 353  
 422 -----EHNGSDLGEVDVGRSPMGHNPACGNASANOQAPASTPLD----- 461  
 354 CHTNSSNNNNNNNNNNNNSSNNNTSSGKTSASNGGTASSGCTADSVAAVDDDDDD 413  
 462 -----ANVLLHTLMLAAGICAMPKLDGTQVDFIKGLLVANSNGIMNEGILNLSAS 514  
 414 DGHGHHQKQWV-----DRLLEQVDVEEDLDVVDVVPATAMARKIGIQR--A 459  
 515 QENSNGNNSLLQQCQCHQHQQHQQHQQCQQAAYARHRLPKSETPTNSSLPNDASE 574  
 460 HON-----LQRLHTTHHQAQ--QHSHNHPSQHNNHPHQHHTHS--DDEQAM- 506  
 575 DPILKIPFKVSGPSSSSSLPGLVGCHNPLNNNSLSISNNNSNHRNGSRSP 634  
 507 -PVL-----AKSEILD-----DYDEMDLEDDDEADNSNDGLNKK- 543  
 635 HSASPMIAAAYAOG---YSAGNSLL-----TSSSSSTQKMMASNIQROINQSG 681  
 544 -----MGSGAGGGGGVDTSTGLPLPLTIPSSAAAAAAMAEQGSTPAMSA 597  
 682 QESLRNGNSDSCSSNNNGSSSLGKKPSISVAKIIGTDTSTRFAS-----PVL 731  
 598 AQAAGAG-ASDLSTIGCGCG-----RPSRSRSDGGGNDGCGGASSSSLSPTGVANL 650  
 732 -----SOCHSAHHLTHQOQQOQLSAQEAIGKGRPKRGKRYNRDPSLV 776  
 651 PVQGVPLSLKKEIDSSDNNSSH--SRHMOORSASAGGAG--GLDYRASHESSTSES-- 705  
 777 EAVRAVGEMSVARAGSYGVPHSTLEYKVKERHLNRPKRREPKPODLVGLTGPAK 836







```

QY 230 TSMDSMAQLAOFSLADFNILNSLASQOQQOQQOQIASAATPTTEVSAAAIAPALMDT 289
DB 291 STTESOPSTTOEBSSTTEGPLSTESSTAVIDQSSSTESODSTTOEBSSTTEGPLSTES 350
QY 290 PSPVDAPLDLSSKPSNSISGVKSVACATPPSGRAYSEEDLSRALQDVANKLD 349
DB 351 ---STEATNESSSTESODSTTOEBSSTTEGPLSTESSTEATNESSSTESOD----- 400
QY 350 ARKASQHEORSILDNRLFKKHHDOHDGDELEDSNDAAEVDSNASTPYPAEF 409
DB 401 ---STTOEBSSTTEGPL--STESSTEATNESSSTESODSTTOEBSSTTEGPLSTES 453
QY 410 ARAQLRLSHLSEHNGSLGEDVDRGSPKMGHRPACGNASANOCAPIPIADVNLHT- 468
DB 454 TEATNESSSTESODSTTOEBSSTTEPL-STEP--STEANESSSTESODSTTOEBS 509
QY 469 ---LMLAAGAMPKLDFTQVGFPIKGLLVANSGIMNEGILNLSAQOENGNAS 523
DB 510 STEDELSTESSTEATNESSSTESOD-----STTOEBSSTTEGPLS-TESSSTES 563
QY 524 LLLQOQHQQHQQHQQHQQOQQOQOQVAAVRRHLPKSETPTNSSLDPN-----ASD 575
DB 564 TESSODSTTOKSSSTESPLSTE-----PSTEANESSSTESODSTTOEBSSTEG 614
QY 576 PILKPSFKVSGPAS-----SSLSRPGLVGHHHPINN-----NSLSISNNS 619
DB 615 PLSTEPSTEANESSSTESODSTTOEBSSTEG-----PLSTESSTEANESSSTES 666
QY 620 NHHSSNHNHNGSNRSPHSASPMILAAVAGGYAGNLSLTSSSSSIOKMASTIQOINQO 679
DB 667 QDSTTOEBSSTESPLSTEPSTEANESSSTESODSTTOEBSSTTEGPLSTEPSTEANES 726
QY 660 SGOELIRNGNVSDCSNNGSSSLGKPKPSISVAKITGDTSRGAAPNLISQOHHAN 739
DB 727 SSTEODSTTOEBSSTEG-----PLSTESSTEANESSSTESODSTTOEBSSTES 776
QY 740 HLTHQOQQOQLSAQALGKTRPKRGKRYNRYDRDLVEAVKAVQGENSVHRAQSYGVP 799
DB 777 ---TEPLSTESSTEATNESSSTESODSTTOEBSSTTEGPLSTESSTESSTESSTESQ 834
QY 800 HSTLEYKVERHLMRRKREPKRPQDVLGTLPANKLOLDKLAGPHG-SKLSNALNQ 858
DB 835 DSTOESSSTTE--SPLSTEPSTEANESSSTESODSTTOEBSSTTEGPLSTESSTEANE 892
QY 859 NNQAAAAAATAAATPNCGLPLPEAGPALS-----QPNMFV 901
DB 893 SSTEODSTTOEBSSTTEGPLSTESSTEGSNESSTESODSTTOEBSSTESPLSTE 952
QY 902 POTNATNAYGLDFNRTTEAMR-----NPOASNHHGLMKSAQDMVENYVDGI 948
DB 953 PSTEANESSSTESODSTTOEBSSTTEGPLSTESSTEANESSSTESODST----- 1004
QY 949 RKTLOASBGNCSAAGNNGSNNGHGHGHGHALLDQLLVKTLPLFTNHRNDYAT 1008
DB 1005 ---TQSSSSTTEPLSTESSTE-----VTQPSPTESLPSNSTGCT 1042
QY 1009 -CSASGSEVAKSGSPMGNYADIKRELSADSGSDEBHSASH-----INNNSDLAHN 1062
DB 1043 PCTTNPSPLEPSPSTPGN-----DDDSNGSGENGNSTSGSPCTTNPDPRESS 1093
QY 1063 KKKSGGGGGGNGGCTNGNGRSTRMTSRDSETDASSPK-----SGENG----- 1106
DB 1094 STPGNDDSGNNGSGENGNSTSGSPCTTNPSPDPBSSSTPGNDDSGNNGSGESGIST 1153
QY 1107 -----GQOHHKMDLNGSGSSSHIKCESEA-----TGHHSP 1139
DB 1154 TGAPYTTNPAQBPSPAPENPDGSGNSSSPBEGATPCTPNAKSTTSYAHATP 1213
QY 1140 GHHT-----TSLHEKLAQIABEYDQAD 1163
DB 1214 KYTTEGNKAETSTLKSPTGTTTGHQEDRDT 1243

```

```

RESULT 13
ID 08SY55 PRELIMINARY; PRT; 1514 AA.
AC 08SY55;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE GH09355P.
GN CG6004.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Cealniker S.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY075323; AL68190.1; -
DR FlyBase; FBgn0036203; CG6004.
DR InterPro; IPR025577; Chitin_bind_Pera.
DR Pfam; PF01607; CBM_14; 4.
DR SMART; SM00494; Ched2; 4.
SQ SEQUENCE 1514 AA; 157140 MW; 1PFC4B064105ADS CRC64;

```

Query Match 5.1%; Score 305; DB 5; Length 1514;

Best Local Similarity 18.1%; Pred. No. 8.5E-09;

Matches 233; Conservative 193; Mismatches 604; Indels 260; Gaps 37;

```

QY 25 HYQ-----DKLTCSHNLNIEOQPIAIGSEDEPEQ-YNHSKEISQSNPNHCKTENH 75
DB 133 HFQFINAGVIEPLSQNLMDKXKDISSSDSPEDATYSQTQVST----- 182
QY 76 RLEQNGSGQLLEEDSENNOTSHD-----SRTPTPGATSPSPPEP 119
DB 183 QVPDASSAESIQSTTQGSRSYDILSTEASLDIILSESIPTPSSSTIILSSSTEG 242
QY 120 IDMPRAKNCVCVGR--LVNAGKLVASATATSSNSNHNHGHDSNSASLP 177
DB 243 -SWESHISTDSISQKVESLLEALYSLQES--SSSESVPVSNPSTGATDDSSSTESLP 300
QY 178 HHSSSSSSNNNSGGRARHIA--AASARATPAATPPANSLEL-----YKLTQRAKM 229
DB 301 DSTOESSSSSPVSFELSTATNESSSESLPRSSSTODSSSTETSFQETSTTDADDES 360
QY 230 TSMDSMAQLAOFSLADFNILNSLASQOQQOQQOQIASAATPTTEVSAAAIAPALMDT 289
DB 361 STTESOPSTTOEBSSTTEGPLSTESSTAVIDQSSSTESODSTTOEBSSTTEGPLSTES 420
QY 290 PSPVDAPLDLSSKPSNSISGVKSVACATPPSGRAYSEEDLSRALQDVANKLD 349
DB 421 ---STEATNESSSTESODSTTOEBSSTTEGPLSTESSTEATNESSSTESOD----- 470
QY 350 ARKASQHEORSILDNRLFKKHHDOHDGDELEDSNDAAEVDSNASTPYPAEF 409
DB 471 ---STTOEBSSTTEGPL--STESSTEATNESSSTESODSTTOEBSSTTEGPLSTES 523
QY 410 ARAQLRLSHLSEHNGSLGEDVDRGSPKMGHRPACGNASANOCAPIPIADVNLHT- 468
DB 524 TEATNESSSTESODSTTOEBSSTTEPL-STEP--STEANESSSTESODSTTOEBS 579
QY 469 ---LMLAAGAMPKLDFTQVGFPIKGLLVANSGIMNEGILNLSAQOENGNAS 523
DB 560 STEDELSTESSTEATNESSSTESOD-----STTOEBSSTTEGPLS-TESSSTES 633
QY 524 LLLQOQHQQHQQHQQHQQOQQOQOQVAAVRRHLPKSETPTNSSLDPN-----ASD 575

```

RA Brandon C.C., Rogers Y.H., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abil J.F., Agdayani A., An H.J., Andrews-Pramkoen C., Baldwin D.,  
RA Ballew R.M., Basu P.V., Bernalde B.J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Beeson P., Berman E., Bhandari D., Bolshakov S.,  
RA Borkova D., Bocchan M.R., Bouck J., Brokstein P., Broctier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA De Palos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz A.C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegam C.,  
RA Jalali M., Kalish F., Kapten G.H., Ke Z., Kennison J.A., Kechum K.A.,  
RA Kammel B.E., Kodita C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mekulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacelle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Sanders R.D., Scheefel F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svayrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodagat, Woxley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C. ;  
RT "The genome sequence of *Drosophila melanogaster*." ;  
RL Science 287:2185-2195(2000).  
[2]  
RP SEQUENCE FROM N.A.  
RA Celinker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Goehayne J.D., Ananitsides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Doreett V., Doup L.E., Doyle C., Dresnek D., Fattan D.,  
RA Ferreira S., Friese E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegam C., Jalali M., Kruse D., Li P., Matei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacelle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouaneavong S., Pittman G.S., Puri V., Richards S., Scheefel F.,  
RA Stapleton M., Strong R., Svayrtkas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M. ;  
RT "Sequencing of *Drosophila melanogaster* genome." ;  
RL Submitted (MAR-2000) to the EMBL/genbank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochnick S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Friese E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E. ;  
RA "Annotation of *Drosophila melanogaster* genome." ;  
RT Submitted (MAR-2000) to the EMBL/genbank/DBJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J. ;  
RL Submitted (MAR-2000) to the EMBL/genbank/DBJ databases.  
[5]  
RP SEQUENCE FROM N.A.  
RA FlyBase ;  
RL Submitted (SEP-2002) to the EMBL/genbank/DBJ databases.  
DR EMBL, AEO03711, AA013691.1 ; -  
SQ SEQUENCE 1249 AA; 133783 MW; 83CDAD995A7DB7DB CRC64 ;



A:Cross-references: EMBL:X90986; NID:g1149498; PIDN:CAA62474.1; PID:g1149500  
 A:Accession: S66150  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 'MG', 428-1085 <WE2>  
 A:Cross-references: EMBL:X90986; NID:g1149498; PIDN:CAA62475.1; PID:g1149501  
 R:Horowitz, H.; Berg, C. A.  
 Development 122, 1859-1871, 1996  
 A>Title: The Drosophila pipequeak gene encodes a nuclear BTB-domain-containing protein  
 A:Reference number: 222972; MUID:96232300; PMID:8674425  
 A:Accession: T45461  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-355, 'E', 357-1005, 'H', 1007-1020, 'Q', 1021-1061, 'ERS', <HOR>  
 A:Cross-references: EMBL:U48358; NID:g1203906; PIDN:AAC47153.1; PID:g1203907  
 A:Experimental source: tissue type ovarian  
 C:Genetics:  
 A:Gene: pipequeak; peg  
 A:Map position: 11  
 A:Introns: 427/3  
 C:Function:  
 A:Description: required for establishing polarity of the developing egg chamber  
 A:Superfamily: BRCore-2 protein; POZ domain homology  
 A:Domain: -123/Domain: POZ domain homology <POZ>

Query Match 4.4%; Score 263; DB 2; Length 1085;  
 Best Local Similarity 17.9%; Pred. No. 1.2e-05;

Matches 216; Conservative 135; Mismatches 398; Indels 458; Gaps 44;

QY 51 EPSQVNSHSEKISQSNPNHCKTENHLEQOH-----NGS-----QLLEE 89  
 DB 166 EQQQHHQQQQQQQQQAEQHQQQQVIAQQQQQQQINAAALTOHGVSSGSVLSQQLISS 225  
 QY 90 EDSENNQOT--SHDSRT-----PTP-----GATSTPSPPEP 119  
 DB 226 SASGSSGSLAGGQAGQTPSGLOPTPRKSLKSKSPULSSGCGAGSGSSGSGTQOP 285  
 QY 120 IDMRPSACNFCVNR-----LITVNAQGLVASAATATSS----- 156  
 DB 286 QPAHHNHQOTLLIQGNPNISIVLSOOTADGNYIPVSGGSDSDADHEHNNHGHG 345  
 QY 157 -----STNSHIHQHDSNSASLPHHISSSSSNNNSGNRAHIAAASATPA 208  
 DB 346 GHTHGHGSHDHSHEHHEKPKICKTEHVAASPASSSSASNN-----AAGVS 395  
 QY 209 AATPANSLELYKLTORAAK---WTSMD-SMAAQLAQS-LADFNLIINSLASQQQQQ 262  
 DB 396 GVTSTGQAIVTQIVVARQDKTKMTSLGKMGNGLLGVPMGFLDFT-----P 443  
 QY 263 QQQIQAATPTTSEVSAALSPALQDTPSPVDAFLDSS-----KSPNSISGD 313  
 DB 444 EPPASATPPTVTEHVDLSCNPSTDRDLSNTEPRLDIDNHLAQIIRLDOSPMSHIS-- 501  
 QY 314 VKSVAACATPTPSGRAYSEEDLSRALQDVVANKLDARKASQHEQRSILDNRLFKMK 373  
 DB 502 -----HHHTGDESNLVQHIKSEVEIEAKGLAAOH-----ALSOAQOQAH 543  
 QY 374 HDQEDHDGDELEDSDNAEAEDVSNASTPVYPAEFARAQLKLSLSEHNSDGLGEDVD 433  
 DB 544 HQAHQOHQ-----QHQQQHHQQQ----- 562  
 QY 434 RQSPKMGHPRACGNASANOAGAPASTPLDANVLLHTLMLAAGIGAMPKLDDETQVDFIKG 493  
 DB 563 -----QH-----LHAQQLA----- 572  
 QY 494 LLVANSGLIMNEGLNLASQENSGNASLLQQQHQHQQHQQHQQHQQHQAAR- 552  
 DB 573 -----OSLQOQQQQQQQQQQQQQQQQQQQQQQQAAAAA 604  
 QY 553 -----HRLPKSETPETNSSLDPNDASEDPILKIPSFKVSQSPASSSSLSFGLVGHH 604  
 DB 605 GVHGOHGHVTHADIGATVMEIDPSQIIGHBGMIIIT-----PEIVNMMS-----SGHM 653

QY 605 HPLNNM-----NSLSISNNNSHSSNHRNGSNRSPHAS-----PMLAAVAQG 648  
 DB 654 DMTNSDTSSEDSMTLANSPI-----DOKEPHYTNLDQHQGLGSGVCPGEGAGCGG 705  
 QY 649 GYSAGNSLLTSSSSSIQKMAANIQRQINQSGOESLRNGVSDCSNNGSSSLGYKRP 708  
 DB 706 GMSGAG-----SGGEKQGFNGPKAMTQDDNMSALDLNKNM----- 744  
 QY 709 SISYAKIIGTDTSRFGASPNLISQCHSHNLTHQOQQQQLSAQELGKGTAPKRGK 768  
 DB 745 SLTKASAI-----YGIPTTLMORAH-----RMGIETPKKEGGTK 779  
 QY 769 NYRDSLVEAVKAVQGEVMSVRAGSYGVPHSTLEYKVKERHLMRPKREPKRPQDLVG 828  
 DB 780 SMNEDALQNLALRSGQISANKASKAFGLPSTL-YKIRREGIR----- 824  
 QY 829 LITGPAN-----KLQDLKAGPRHGSKLNALKNQNNQAAA-AAAAAAAAA 875  
 DB 825 LAAPFNAPPTTWPRDEDLRALALRAG-----NTSVQKASAEFGIPGTLYGR 872  
 QY 876 ATPNGKLPLFPAQPOLSFQPMWFPQTNATNAYGLDFRITBAMKNPQASNHGLMKS 935  
 DB 873 CKREGIELSRSNPTP-----WSEDAMNEA--LNSVRVQMSIN-QAAIHYNLPS 919  
 QY 936 A-----QDWENVYDGIIRKTLQASEGNSAAGNSGNSGNHGHG---HA 983  
 DB 920 SLVRFKRGKDYVANT-----SGVALLNTGNTTGSLEIIEHSGENSLHM 965  
 QY 984 LLDQLVKKTPLPPTNHNNDYAATCSASGESYKRGSP---MGNYADIKRERLSADS 1039  
 DB 966 LQQGFPTSPSPHPTPQH-----STPQHSSAQGGPPTPQMQQYVVHMQQQQQQQ 1018  
 QY 1040 GGSDEHSHASHIN-----NNSDLANKNK-----SGGGGGGGGNGCTN 1079  
 DB 1019 QQSEPHHHPQHMQQDVVATSSQVHSGQQQQQLQIYOHNGTPVTASGSSAVSGTSKRX 1078  
 QY 1080 GNGRSSR 1086  
 DB 1079 KANKOKR 1085

Search completed: October 29, 2003, 12:31:30  
 Job time : 64 secs

QY 791 RAGSYGVPHSTLEYKVERHLMRPKREPKQPDVLGTLGPANKLQDLKLXAG--FHGG 848  
 DB 904 LAGSTSGSSSTYGVSSDSR-----DGSVSTSGSSSNTDASTDLAGSSTSGG 950  
 QY 849 SKLSNALKNQNNQAAAAAATAATPNGIKLPLFAGPQALSFQPMFMPTQTNAN 908  
 DB 951 SSTYGVSSSDRGSVSTSSSSNTDASTDLTG-----SSTYS 986  
 QY 909 AVGLDENRITLQAMRNQANHHGLMKASQDMWENVYDGIIRKTLQASENGSAGNGSN- 967  
 DB 987 G-----GSSYGVSSDSRD-----GSVSTSGSSNT 1012  
 QY 968 -----GSGNGHGHGHGHALLDQLLVKTPLEFTNRRNDYATCSSA-----SG 1014  
 DB 1013 DASTDLAGSSTSGSSSTYGVSS-----SSNRDGSVATGSSSNTDASTTE 1056  
 QY 1015 ESVKSGSPMGVADIKRE-KLSADSGSSDEEHSASHINNNSDLA-----NKXK 1065  
 DB 1057 ESTTSGSSTEGSSSSHDGVSITDGSSTSGASSSASTAKSDAASEDEDFWMMNRKX 1116  
 QY 1066 SGGGGGGGNGGQTNNGNGRFSRMTSRDSETDASFSGENGGQGNHKKMDLNGSSSSSH 1125  
 DB 1117 S-----GSGHKATVGSSTTDKTSIDASSTDSSTSGASTTSGSS--STSGGSSTSDA 1170  
 QY 1126 IKCESEAAATGHSRPGHHTTSLHE 1149  
 DB 1171 SSTSSSVSRHHS---GVNRLHK 1191

## RESULT 14

F90073  
 Hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #ext\_change 22-Oct-2001  
 C:Accession: F90073  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
 ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
 C.; Shiba, T.; Hatori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: AB9758; MIMD:21311952; PMID:11418146  
 A:Accession: F90073  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2271 <KIR>  
 A:Cross-references: GB:BA00018; PID:g13702612; PIDN:BA943752.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 A:Genetics:  
 A:Gene: SA2447

Query Match 4.5%; Score 269; DB 2; Length 2271;  
 Best Local Similarity 17.2%; Pred. No. 1.5e-05;  
 Matches 195; Conservative 200; Mismatches 506; Indels 236; Gaps 27;

QY 43 IAIAGS---EDPPOYNNHSSKEISQSNPNHCKTENHRLQCHNGSOLLEEDSENNTQSH 99  
 DB 1015 ISTSGSLASDSKMSVSSSMSTSGGT-----SESLSDQSTSDSDSKLSLS- 1064  
 QY 100 DSSRPTTGATSTPPEPPIDMRPSAKCNFCVNGRLTVNAQKLVASAAATATSSSTS 159  
 DB 1065 ---TSOGSSTSTST-----STSAVRTSSQSTSGSMSASQSDSMSTSTSFDS 1110  
 QY 160 NSHHQHSDSDSSASLPHHTSSSSSSNNNSGNAARHAAASAAATPAATPANSLELY 219  
 DB 1111 TS---DSKASSTAS-SESIISQASSTISGVSSTSTSLSTNSERTSTSVSDSTSLSTS 1164  
 QY 220 KLTGTAAMTSMDSMAAQLAQSLLADPNLINSASQOQQOQQOQIASAVPTTSEVA 279  
 DB 1165 ESDSISESTSTSDISEAISSESTISISESNSTSDSESOASAFLSISESTSEST 1224  
 QY 280 AAI-----SPALKDTPSPVDAPLDLSKSPNSISGDVKSVRACATPTPSGRRAVSE 333

DB 1225 ESVSSSTSESTSLSDSTSESGSTSLSNSTSGSASIS-----TSTSIESTSTPKS 1276  
 QY 334 EDLSRALDQVANKLDARKSAS---QHHEQSSILNRLFKKHHDQEDHDDELED--- 387  
 DB 1277 ESVSTSLSNSTSLSNSTSLSTSLSDSTSDSKSLSTSNSTSDSTSTSTSTSTST 1336  
 QY 388 -SNDAAEAEVDSNASTPYPAEFARQRLKLSLSEHNGSDLGEDVDKSPKMGHNPACG 446  
 DB 1337 LSGSTSESESDSTSS-----SESKSDSTSMISMSQSTSGSTSTST 1377  
 QY 447 NASANQAPASIPLDANVLLTMLAAGIAMPKLDFTQVGDPIKGLLVANGGIMNEG 506  
 DB 1378 STSLSDSTSTSLSTSLAS-----NMQS 1398  
 QY 507 LNLTLASQENNGNASILLQOQHQQHQQHQQOQQOQVAAVRRH-LPKSETPENS 565  
 DB 1399 GVDNSASQASQANSTST-----STSESQSTSTYTSQSTQSGESTSTST 1443  
 QY 566 SLDPNDASEDPIKIPSFVSGPAPASSSLSPQGLVGHHPPLNNNSLSISNNSHSSNS 625  
 DB 1444 SLSDS-----TSISKSTSGGSTSTASLSGSESESDSQSISTSAESTSEASST 1493  
 QY 626 HRNGNRPBHASPEWLAALAAVAGGYSAQNSLITSSSSSIQKMAANIQRIQNEQSGESL 685  
 DB 1494 SLSDSTSTSNQSGASTSTSLSN--SAGASESDSSTSLSDSTASMSQSESDSQSTAS 1550  
 QY 686 RGNVSDCSNNQSSSLGYKKPSISVAKIIGDTDSFGASPNLISQGHSHAHLLTHQ 745  
 DB 1551 LSDSLSTSTSNR-----MTSLASTSVTSGSTSESTSESTSTSTSLSDSQ 1599  
 QY 746 Q-QOQLSAQEAALGKTRPKRGKRYRNYDRDSLVEAVKAVQGEHVRAGSYGVPHSTLE 804  
 DB 1600 STSRSTSGASGASTST-----STSDRSTSGASTSTMR 1632  
 QY 805 YKVERHLMRPKREPKQPDVLGTLGPANKLQDLKAGHGSGKSLNALKNQNNQAA 864  
 DB 1633 TSTSDSQSMSTSTSTSTMSDSTSLSDSDSTSDSTSTASTSGMSVSTSLSDSTSTST 1692  
 QY 865 AAAAAAATAATPNGIKLPLFAGPQALSFQPMFMPTQTNATYAGLDFRITAMNPNP 924  
 DB 1693 ASEVMSASISQSNSESVNDSESVSESNSSDSK--SMSGSTSVSDGSLSVSTSLKS 1750  
 QY 925 QASNHGHLMKSAQDMENV-----YDGIIRKTLQASENGSAGNGSNGNGHG 975  
 DB 1751 ESVSESSSLSGSQMSDVSSTSDSSLSVSTSLASSEVSDDSLSDSKSTSGS----- 1804  
 QY 976 HGHHGHALLDQLLVKTPLEFTNRRNDYATCSSAG-ESVKSQSPMGVADIKRE 1034  
 DB 1805 -----TSTSGSLSTSTSLSGSEVSESTSLSDSIS----- 1836  
 QY 1035 LSADSGSSDEEHSASHIN-----NNNSDLANKKSGGGGGGNGGQTNNGNGRSM 1087  
 DB 1837 -MSDSTSTSDSDSLSGSTSLSTSLSDSLSDSKSLSSQSNMGSE-STSTSVSDSQS 1894  
 QY 1088 TSPDSETDASFPKSGENGGQGNHKKMDLNGSSSSSHIKCESEAAATGHSRPGHHTT 1144  
 DB 1895 STSNSTQPTDSNISASBDSMSSTSDSNISGNSNTSTSLS-TSDSMGGSVSVSTST 1950

## RESULT 15

S66149  
 gene p13queak protein A long form - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #ext\_change 23-Sep-2002  
 C:Accession: S66149; S66150; T45461  
 R:Weber, U.; Siegel, V.; Mlodzik, M.  
 A:Title: p13queak encodes a novel nuclear protein required downstream of seven-up for t.  
 A:Reference number: S66148; MIMD:96134923; PMID:8557044  
 A:Accession: S66149  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1085 <WEB>

Dd		435	ETVANIIGLKRLPEOLQOOSAS--NPBOKSYRNAPSMPMSGSSSTNGSSSLGTADTASTS---	490
Oy		343	VVANKLDAKASAOHNEORSILDNBLFKYKHDDODEHDHDELE-----DSNDDBAEAYD	397
Dd		491	-----TSTSSHRRK-----KKGSKEPRDAAGKKRKLUAETISSQDGMKYK	533
Oy		398	SNA\$TPVVPAEFAA-----QARKLSHEHNHSGDLGEDVDKSPKMGHPACGNAS	449
Dd		534	ITAK-PNKLDPKRSHSLASGELDOQL---KLDTSTSEALNR---TLGEBARINSJ.	585
Oy		450	ANOGAPASIPLDANVLHLTMLAAGICAMPKDETQTVDFIKGLLVANSGGIMNEGILN	509
Dd		586	VVGGAFTPTPTT-----AEPEEQOOQQO-----	608
Oy		510	LISASQENSNGNASILLLOOQHQHQNHOHQOQOQOQOHAAYVRHRLPKSE--TPETNSSJ.	567
Dd		609	-----OQOQOPOOQOQOQOQOQOQOQFVV-----LPKIXDLTTLPTSPL.	646
Oy		568	DPNDASEPILKIPIFYXVGRPASSSSLSGVLGCH-----	603
Dd		647	PPS-----LFKAYTP-STTPTAHTYAAGGPKKOQOQOMPOOPVALQOSMAKTN	694
		604	--HHPLANNNSLSISNN\$NH---SNSHRGNSNSPSHASPYMLAAVAOGGSAGNSLL.	657
		695	PAKPRLSSNNNR--KPN\$GHFAVPAPRTIRMYTHMORYOST\$SLA-----SAANKRP	745
		658	TSSSS-----SIQKMASNIQRIENGSGQES-----LRNGVSDC	693
Dd		746	KRSM\$LD\$HPAKQARLSQAQMAMASSYAKLHMOTSNQAKSOAAFLPMPQR\$YGLPDL	805
Oy		694	\$SNNGSSSLGYKXRKSISVAKIIGTPTSRFGASPRLYSQOHNSAHNLHNOOQOOLSNO	753
Dd		806	-----GSKPTLPMLCBAS\$SSQV---TTTPRAPRTBSITYFS\$EBNIPVALLEYRLRVNQ	858
Oy		754	EALKG-\$TRPKRGKYRYNDRDSLVEAVKAUVORGENSVHRAGSYGYVP-HSTLEYKYKERH	811
Dd		859	SAGCGGLTMPRLS-----PATTSARLMGPRLPALPKAHGHC\$AKRS	901
Oy		812	LMRPKREBKP-----OPDLVGLTGPRANKLOLDKLAQBHG\$SKLSN--	853
Dd		902	COMPTMPMPLPLPLMPMTTIPAIKSPPL\$VALSGORN-----KGNSSNSMAY	951
Oy		854	-----ALKONUNQAIAAAAAAAAAAAAAATPNGKLPLF\$AGFOALSFGQNMFPQMNATN	908
Dd		952	RTPSPRALINLRNTAAPHSF\$KSP\$KYEANSKK\$P-ALGCG-----KMGRA	100
Oy		909	AYGLDFNBITEAMRNPQASNHHGLKMSAQDWENTVYDGIIIRKTLQASEGNSAANG\$NG	968
Dd		1001	A-LDK\$KTS-----LREP\$PAVO\$AV---TATAT\$VTTTAAGA\$A--	103
		969	SNGNGHGHGHGHALL--LDOLLVKYKTLPTPHNRNDYUATCS\$ASGESVYKRS\$PMG	102
		1037	-----GAGTG\$TALAKDADLIDLISANP-----GRESND-----AKLAPNSPPAG	107
Oy		1026	NYADIKBERLSADSGSSDEEH\$ASHINNNS-----DLAHNKK-----SGCG---G	107
Dd		1077	N-----NNNNNNNNNNNNNNNNNNNNNSTNSLEALNLKIKONISANGSG\$TTS	112
Oy		1071	GQGNGQTNGNG-RSSRMTSRDSETD\$AF\$KSGENGOQNHKMDLNG-GSS\$SSHIKC	112
Dd		1126	GSNSGJTNGDDLQNLHMLS\$SATAREKISIVASA\$G-----N\$SGTSS\$SAXP	117
Oy		1129	ES-----EAATHGHS\$CHHTTSLIHLEKLAQIKAO	1158
Dd		1176	KNAALVBPONASVSIPNPSALA\$FRNOPAASISKPLVTRAEE	1223

RESULT 13  
S52714  
bericlinB - silkworm  
C:Species: Bombyx mori (silkworm)  
C:Date: 19-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Sep-1997  
C:Accession: S52714

R.Garel, A.A.; release, G.G.; Prudhomme, J.J.  
submitted to the EMBL Data Library, March 1995  
A.Description: Structure and organisation of the Bombyx mori sericin I gene and of the flanking regions  
A.Reference number: S52714  
A.Accession number: S52714  
A.Status: preliminary  
A.Molecule type: mRNA  
A.Residues: 1-1217 <GAR>  
A.Cross-References: EMBL:Z48802, NID:g7556699, PID:g755700

Query Match	4.5%;	Score 270.5;	DB 2;	Length 1217;
Best Local Similarity	18.5%;	Pred. No. 5.7e-06;		
Matches 238;	Conservative 174;	Mismatches 463;	Indels 389;	Gaps 45;

	Query Match	4.5%	Score 270.5	DB 2	Length 1217	
	Beet Local Similarity	18.5%	Pred. No. 5.7e-06			
	Matches	238	Conservative 174	Mismatches 483	Indels 389	Gaps 45
Oy	5 SYEISLEAEVEECGRGROWKHYODKLTCSHLNEHOOPAIAGSEDEPSQ-----YN	56				
Db	158 SYGVASADSGSASSRRRQANYSDX-----DITAAKDPRSADSSRRSMAYVN	205				
Oy	57 H----SSKEISQSNNPHCKTEN--HRLBOHNG-----SOLEEFDESENQT----	97				
Db	206 RDSDDSEFAGLSDRSASSSKNDNFVYRTKSIGOAKSSRSSSHQEEDAYNNSPDGSY	265				
Oy	98 ---SHDSRTPTPATSTPSPPEPIDMRPAKCNFCVNGLLTYNNOCKLVAESAPAT	154				
Db	266 NAGTRDSSSTSKKKASTTVADKQI-----PAANDRSSKKOLKOSSAITS	311				
Oy	155 S-----SSTSNIHHOHSDSNSA-----SLPHHSISSSSNNNSGNGRAHI	198				
Db	312 SGPKTSTVSSKDRQYSNDKRKSDAYVORDGTVAYSNKDSKTSRGSTNTNVADONSVPD	371				
Oy	199 AAASARATP-----AAATPANSELKYLTORAAKMTSMDSMAQLAQFLSA	246				
Db	372 SAASPOTSXYDRGSDKNIVIAHSSGSGSONOKSSSYRADKDGFSSTNTEKSFVS---	428				
Oy	247 DENLINSLASQQQQQQQIASATPTTSEVSAAALIPALDTPSPVDALUDJSSKRP	306				
Db	429 -----SSNSVETISDGASARESSA-BDT-----KSS	454				
Oy	307 NSSIGDVSVACATPPSGRRAYSIEDL-----SRALQD-----VVANKLDARKASQIH	357				
Db	455 NSNWGD-----FTGEHEELPDVVSYOKIEGKPIIMKVLPVESKASQS	499				
Oy	358 HEQRSLDNRLFMMKHDOEQHDQDELEDSS-----NDAAEA-----VDSNASTP	403				
Db	500 SSRSQSGASAYSSSSSSSTLSEBSEVDIDLGILGMWMSNDKRAQAAGATKXEASS	559				
Oy	404 VYPAFAPAQO-----LRKLHLSEHNSDLGEDVDROGSPKMGHRPACGNABAN	451				
Db	560 TQATTVSAGDADSASYTMWNPRRSHSSSSSSASSSSGNNVGSSGSSGOSTGCMNAGH	619				
Oy	452 OGAPASILEDNLVTLHTMLAAGICAMPKLBDETQVGFIFGLVAVANSQG--IMNEGILNLT	510				
Db	620 LGTVSSTGSTNT--DSSSKSAG-----SRTSGGTSTYGSSHHGGSVYSTGSSSN	669				
Oy	511 LSAEOENGNASILLLOOQHOOHQHHOQHOOOQOHHAAVVRHRLPKSETPETNSSLDPN	570				
Db	670 TDSSTKMAGSSSTSGSTSTGYSSSR-----GGSVSTGSSSNTD	709				
Oy	571 DASEDPILKIPIFKVYGPASSSSLSPGCLVGCHHPLLNNNNSLSINNHSNHSNRHG	630				
Db	710 S-----STXSAGSTSGGTSTGYSSRH-----RGGSVSTGSSSNTDSTKNA	754				
Oy	631 NRSPIHASPMLAANAOGG--YSAQNSLTLTSSSSLOKMASNSIQOROINQESGOELNKG	688				
Db	755 SRTSGCTSTGYSSSHRGGSVSTGSSSNTDST--KNAGSRTSGGTSTGYSSHHGGC	811				
Oy	689 NVS-----DCSSNN-----GGSSSLGYKKDISVAKIIIGTDTSRFGASPULLSQO	734				
Db	812 SVSSTGSSSNTDSTKNAKGSRTSGGTSTGYSSSH-----RGGSVSTGSSSNTDST	864				
Oy	735 HHSAHHLTHQOQOQOLAQELGKTRKRGK---YNYNRDLSLYEVAVRGEMGVH	790				
Db	865 KNA-----GSSTGSSSTGYGSSDSDRDGSVSTGSSSNTDASTD	903				







[illegible]

OY	693	CSNNNGSSSLGYKKPIS-VAKITIGDTSPFGASPYLLSQDNHNSAHNLTHTQQOOOLLS	751
Dd	674	ASSNNNTSGSIN---NISIGNMNGQDSR-----NLKRVSSAPRMONTL-	715
OY	752	AQEALGKGTREPKEGKYRYNYDRDSLVEAYKA--VORGESYHRAGSYYGVPHSTLEYKVK	808
Dd	716	--AVVNGPP-----PLPEPLRLAPYOQRNMLSTDOPNS-----	748
OY	809	ERHLMPREKREPPODLVLGTGPANKQLDRLTKXGPHGSKLSNALKNONNOAAA--	865
Dd	749	-----MTGPT--MHLOKNVGPNQPR-----TNANGTGAALPPP	780
OY	866	--AAAAAATAATPGLKLPLEAGPOALSFOPMFMFOTNATNAVGLDFNRITTEAMRN	923
Dd	781	VPAEPNPVATEVAVDSDGLEV-----VEEPSILRPSELVRGNHNRTWSTISANKAKALL-N	833
OY	924	PQASNHHGLMKKSQODMVENVYDGIIIRKTLQASEGSGAA-----GNSNSKSNNGH	974
Dd	834	AGSTNGSSTLAASDD-SQRYGGSYH-----AANGSAANNGHPCYSGNKGNGQSMAANGN	888

```

Db      887  GNGNGGGG-----NYHSHPNDPHAT-----GQPHQYTSPLYG-----  918
OY      1035  LSADSGGGSDEEH 1047
          |||||
          |||||
Db      920  -----GGSSAEDH 927

RESULT 11
S57237
forked protein 5.6K - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 22-Jun-1999
C:Accession: S57237
R:Hoover, K.K.; Chien, A.J.; Cores, V.G.
Genetics 135, 507-526, 1993
A>Title: Effects of transposable elements on the expression of the forked gene
A:Reference number: S57236; WUID:94063487; PMID:8244011
A:Accession: S57237

```



[illegible]

## RESULT 7

parallel sister chromatids protein - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13349  
R:Gandhi, R.L.; Goldberg, M.L.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: Z17657  
A:Accession: T13349  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1768 <GAN>  
A:Cross-references: EMBL:U40214; NID:g1100982; PID:g1100983; PIDN:AAA91230.1  
C:Genetics:  
A:Gene: pacc  
A:Cross-references: flyBase:FBgn0004655

Query Match	4.7%;	Score 282.5;	DB 2;	Length 1768;
Best Local Similarity	19.4%;	Pred. No. 2.3e-06;		
Matches 238;	Conservative 160;	Mismatches 467;	Indels 361;	Gaps 45

OY	48	SEDESOXNHSKELISQSPNPNCKTENRLEOHNHSGSLTL-EEBDSENNQJSHDSRPPT	106
OY	203	ASEOGEQSSASGAKOKKKPK---EKKLKEAPRPRULGARAAVYREVEDDERPPT	258
OY	107	PGASTPSPRPPPEIDWRPBAKCNFCVNGRLTLVNAOGLVAESAATAT-SSSTSNSHIQ	165
Db	259	P-----TKDLIIPRAGROPAPVAATATATAAASBAFIS	292
OY	166	HDSDNSASLPHHIISSSSNNNSNGNARHIAAASAPATPAATPANSLKYLLTOR	225
Db	223	TFGSGSESLPPLPPTSPABASSTSOBPS---ASGSASNPSPASRTEPHPIVLIRISG	349
OY	226	AAKMTSMDSMAAQLOAFSLILADFNLIINSIASQOOOQOOQOQIASAVTPTTSEVSAAISPA	285
Db	350	TSRLVSTDS---EEPPSSPAHONOLINOLSVTEEBEAEKSGDETPASRPTKT---VKPL	403
OY	286	LKDTSPSPDAPLJLDSKSPNNSISGVKSVACATPPSGPRAVSEEDLRALO---	341
Db	404	RPTLADSVD-----GSSAAVCGAAGSGDSFEERKQSGLEPNDEDEEEDEDEEPPRI	458
OY	342	-----DYANKLDARKASCHHEORSILDNFLFMKCHHDQO	378
Db	459	NYCTVKISBPKPKERKLIITVDIARNAIAAAAA---ESRESEKSKSKKHK---KOL	513

```

QY 379 DHODELEDESDDAEAVDSNASRP-----VPAEAPARQRLKLSLSEHNSDL 428
Db 514 LAAGSGAAPASGATPAELINSEFKTPSHLALSEANSQOAOHTPSHLQHLQHLHFORGSV 573
QY 429 -----GEDVDRGSPKMGRRHPACGNASANOGABA----- 456
Db 574 ISPTRRSDHPDSSSVLGSISSKGNSTPOLLAOVAOEDSCVLRSSSVTSDLETSON 613
QY 457 -----SIFLDANVLIHTML---AGIGAMPKIDETGVVGPFIKLLVANSOGJINEGLN 509
Db 634 SSLVAPSPDIESRLESMMMTTIDGNGTGAASAVETP-----LOEDILA 676
QY 510 LLSASOENSGNASLILLQOOOHOQHOOHOQHOOOHOQVAAVYRRLPKSETPETNSSLDP 569
Db 677 VLRGVEPRUNGNTD-----PEPTEEDQOQOPKATRGGRKA-----NNNVDV 720
QY 570 NDASEDELPLKIPSKVSGGPASSSSLSFGVLGHHHPLNNSLSISNNSHNSNRHG 629
Db 721 TPPTTE---TRRGRAKAGADATTAISP-----PTGKRN---TRGTGSKKAEDEV 765
QY 630 SNRSPHSAPYLAANAQ-----GGYSAGNSLTTSSSSSIQKMAANIQRQI 676
Db 766 DMEVDETAMTVPANEOLEQATLPRRGVRAAARANNNLASVNNNIKTAAN-----L 820
QY 677 NEGGQESLRNGVSDCSSNNGSSSLGYYKKEPSIVAKIIGCTDTSRFGASPNLLSQOHH 736
Db 821 SAKAEASRLAEGVAA-----GGAARSYGRKKRKNQOVVL-----OOE- 858
QY 737 SAHLTLHQOOOQOOLSAOEALGKGRPKRGKYRNRDROSLVEAVKAVORGENSVHAGSY 796
Db 859 -----PVPBEETDAAE---EOPTRAK----- 878
QY 797 GVPHSTLEYKKEHMLPKRKEBKPORDLVGLTGPANKIQLDKLKACP---HGSKL 851
Db 879 -IPIHTD-----HREHPDHDPPDDELSSNNSNSSLQHDSSSSPPEDEKFXKDF 929
QY 852 SNALKNQNQAAAAAATAAATNGUK---LPLFEAGPOLSPFPNMFQOTNATN 908
Db 930 KRTLLDTQGANAGAGAAALAAPESSGEBGRVAKVLTSKKKGSIFSRALVPSDOEQ 989
QY 909 AYGDLFNRTIEMRNPQASNHGMLKMSAQDVENVYDGIIRKTLQASEGNSAANGSNG 968
Db 990 A-----TYAKKH-----LYKHSMD-----AALEANG--GJNSDA 1011
QY 969 SNGNGHGHGHGHALLDOLLVYKTPLPF---TNRNRNDVAAATSSSAGSEVKSAGSF-M 1022
Db 1018 SNASASGVAVGAKDHLHTLAAKSDGDFGSPSNNNGSSSACSSA---STLGDSPAL 1074
QY 1025 GNYADIKERKLSADSG-----GSDSEHSASHINNNSDLAHNKK 1065
Db 1075 G-----KISRLAGQGVATSTSSDAPDLDEPIAGELDLERSAA-----GA 1116
QY 1066 SGGGGGG--GNGQTNNG--RSSRMTSRDSETDASSFKSGENGQGNHKKMDLNGSS 1122
Db 1117 SAGGTGATGGGATGGGGPYRVPRKTK--DYUVPVRVVKTA-----HOIOEIGEYOE 1167
QY 1122 SSSHKCEAATGHSFGHHTTSL 1147
Db 1168 MDDVVEYTLDALQPHNPATRCLSL 1193

```

## RESULT 8

parallel sister chromatids protein - fruit fly (*Drosophila melanogaster*)  
C/Species: *Drosophila melanogaster*  
C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #cont\_change 17-Nov-2000  
C/Accession: J13610  
R/Murphy, L.; Harris, D.; Barrel, B.  
submitted to the EMBL data library, April 1999  
A/Description: Sequencing the distal X chromosome of *Drosophila melanogaster*  
A/Reference number: Z17668  
A/Accession: J13610  
A/Status: preliminary; translated from GB/EMBL/DBD

```

Db 97 000000HRSISNMDRTVNAKFKPASSVAGGPNPNVRRNSMLTSPRGVIGTGCTGNIRK 156
Qy 218 LVKLLTORAKMTSDMAAQLAOFSLADFNLINSLAQ000000-----QOIASAV 270
Db 157 LRV-----SLLTSMNHFAVCYPRPSNIYQNSNNAGSNALQRTTSSRLNMMSKVAA 211
Qy 271 TTTTBEVSAAIISPALKDTPSPSVDAPLDLSK-----PSPNSISGDVKS 316
Db 212 TPTT--VSFRASSNSLSTSTSL-APKSSSSSGGSGNSTPQ000000LVSSNNSSSSNS 268
Qy 317 VACATPTPSGRRAVSEEDLSRALQDVANKLDARKASQOHQEQSILDNRLFKKHQHQ 376
Db 269 FTKASSPNNGARVSGAATSAATGTTA-----AGSHHQPHHHHHHHHHHHQHNNH 321
Qy 377 EQ----- 378
Db 322 000000TSLSGHSLTVAGSGASAGGGGGSSGTAAGCTNRKPKTTSPFITSVT 381
Qy 379 -----DHDGELEDSDNDAAEV-DSNASTP-----VYPAEFARAO 413
Db 382 VGHPRKLNAGDTGDESADDLDESHDNDNRITDLENETPSMEDTFSKEEYVYANNAJST 441
Qy 414 LRKLSLHSHNG----- 431
Db 442 NARVLPITSQYGLVVVDPIAPSLGCTIQNVQVNSDNIIVVSGAVTPGCTKKDKDET 501
Qy 432 VDRGSP-----KMGK-----HPACGNASANOAPASIPLDANVLLHTLWL 471
Db 502 QHRSRFPKVKLESTEPFKGRHMCMDYLDHSSVGNCGNNKTKGSSISEAIAAT-TDGG 560
Qy 472 AAGIGA--MKLDETQTVGDFIKGLVANSQGINNEGILLNLSAQSNSNANSLILQO 528
Db 561 AAGVAGSEAPRAHKTQS-----MLPPTOKINE--NHELANSTANNVNAVEEQOQ 609
Qy 529 00HQH00H----- 549
Db 610 00000000TIVGNALLTKTLPVALRNVSRSSVTRSPNATVEFLSPNLLAQ000000Q 668
Qy 550 AYRHLRPKSETPETNSSLDPNDAE-----DPLIKIPSFVK 585
Db 669 -----FDSVNAHAASSFPNAGDPNNMDYARTAMOLHOTLOOKOREDMAYP---- 716
Qy 566 SGPASSSSLSPGG-LVGGHHHPLNNNSLSISNNSNHSNSH--RNGSNRSPHSAPWL 641
Db 717 PGAGGYANYQNGDSAVGA-----ASNNSAAATGESQLSSTYVQ000000PLSPAPLT 772
Qy 642 A-----AAVAGGYSAGNSLITSSSSSIQKMAASNIQ-----ROINEGSGGSLRNGVS 691
Db 773 FOAAPTFAAVALAG--QSPNFOLEQ000000Q0ATSDGIVPQPPNPO000000TPOQSTA 830
Qy 692 DCSNNGSSSLGYKKPSTISVAKIIGTDTSRFGASPNLLSQCHSAHNLTHQ000000L- 750
Db 831 QOAAAAAATSAVTAAPPOQTSTNTSMAAVTTGOGOTMPLS-----HWTSEQ00PNIG 884
Qy 751 -SAQFALGKTRPKKGYRNYDRDSLVEAVKAVORGEMSVHRAGSYG-----VPHSTL 803
Db 885 AAAAAAAGGTAA-----TSVAPRAIP--TLQASAPSTIADPQGLVWPQ00Q 931
Qy 804 EYKVERILMRP-RGREKRPQDVLGLTGRPAKLOLDKLKAGPHGSGKLSNALKQONNOA 862
Db 932 000H0EEQ00Q00000PLPPANIASASANNNSNLMITNTV-VATGEATNTALLTLTDOA 990
Qy 863 AAAAAA-----AAAATPGLKLPLEAGPQALSFQPNMFQPTNA--TNAVGLDENFT 918
Db 991 TRLAALAAATGAAAAAATCATSAATAATQOQIQLOQ00PNAESTESASGTSAAVAD-NKIE 1049
Qy 919 EAMRNPQASNNHGLKMSAQDWE 941
Db 1050 QAM--DLYKSH--LMIAYREEVE 1068

```

RESULT 6  
S69205

stripe a/b protein - fruit fly (*Drosophila melanogaster*)  
 C1Species: *Drosophila melanogaster*  
 C1Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Jul-2000  
 C1Accession: S69205  
 R1Frommer, G.; Voithnueggen, G.; Paeca, G.; Jueckle, H.; Volk, T.  
 EMBL J. 15, 1642-1649, 1996  
 A1Title: Epidermal egf-like zinc finger protein of *Drosophila* participates in myotube gu  
 A1Reference number: S69205; MUID:96203082; PMID:8612588  
 A1Accession: S69205  
 A1Status: preliminary; nucleic acid sequence not shown  
 A1Molecule type: mRNA  
 A1Residues: 1-1180 <FRO>  
 A1Cross-References: EMBL:U42403; NID:G1147788; PID:AA802355.1; PID:G1147789  
 C1Keywords: alternative splicing

Query Match 4.7%; Score 283.5; DB 2; Length 1180;  
 Best Local Similarity 20.6%; Pred. No. 1.2e-06;  
 Matches 236; Conservative 128; Mismatches 457; Indels 323; Gaps 41;

```

Qy 35 LINEEQPFIAGSEDEPQYNSKKEISQGNPNCKTENHRLQOHNGSOLLEEDSEN 94
Db 3 LTRKQGNEL-IVGQGHSAATSSASASAGATSSPGLSQGNLVA--GAGATNTSTAN 58
Qy 95 NOTSHDSRTPTPGATSTP---SPPEPIDWRPSAKCNFCVNGRLTVNAQGLVAESA 151
Db 59 SODSLNTPPTTLGLSRNPLOFAPPAPPIAVSPSAGPFGYQYOTANAPRLHNSPAA 118
Qy 152 TATSGSTNSHHIHQDSDNSASLPHHISSSSSNNNSGNRARIHAARATPAAT 211
Db 119 TSEVTAPOPPV--ELDEYVDILQVOQLDSSA-----AAAA 154
Qy 212 PANSLLEYKLLTORAKMTSDMAAQLAOFSLADFNLINSLAQ000000QOIASAVT 271
Db 155 AANNPTTEGQVQ00Q00TVQPHHQ00Q-----Q000000PQOVLAKPR 201
Qy 272 PTTSEVSAAIISPALKDTPSPSVDAPLDLSKSP-----NSS 309
Db 202 PRINLOKATEVAQAQAVESSSPCSRVLDYPSRYGPNYHHHTPGEEDLVALMFGSN 261
Qy 310 ISGDVKSARACATTPPSGRRAVSEEDLSRALQDVANKLDA--RKASQNHQESILDN 366
Db 262 GTGGV-----APGTPSAM-----IMEGLETVAAPTHNAFLLTETAAAHFNVLST 309
Qy 367 RLFKAKHHQDQDHDGELEDSDNDAAEAVNSASTPVYPAEFARAOJRLKLSLSEHNS 426
Db 310 CLFK-----TSTAQSSGSPRTATYISPAQFG-----SHGGA 343
Qy 427 DLGEDVDGSPKGRHRA CGNASANOAPASIPLDANVLLH--TLMLAIGAMPKLDDET 484
Db 344 STSSN-----SLNSSSTAASSSTSSSLHYTTASAAAAAASAAAAASAN 386
Qy 485 QTVGDFIKGLVANSQGINNEG-----LNLNLSAQENSGNLSLLOOQ00HQHQHQ 536
Db 387 NSV-----LRAAQATAPPTGSGPGHVAVQSAFASGRSASHLSLNTSGHSPTSS 440
Qy 537 OHHQ000000VAAVYRHLRPKS-ETPETSNSLDPNDASEBDILKIPSKVGPASSSLS 595
Db 441 AVEQVEAHKQILEA-----LPGLDNTPVTTSSDIFSFGPTTVVPRPTGSISEEDLSLE 496
Qy 596 P-----GGLVGGHHHPLNNNS--LSI-----SNNSNHSNHRGCSNPSHSAPWL 641
Db 497 FOIVSVASPVLSHSCSPLEERSTPPALAIYVESSNNSCNMYPOHNNNNNNNNNTTSSST 556
Qy 642 AAAAAGGYSAGNSLITSSSSSIQKMAASNIQROINEGSGESLRNGVAFSCSSNNGSS 701
Db 557 TTTSKQ-----TTSESNTGCVSPGNHTSHQ00-QOOLQHNNTSSNSNCHSH 605
Qy 702 SLGYKKPSTISVAKIIGTDTSRFGASP-----NLISQO-----HSGAHNLTH-- 743
Db 606 000000-----Q00QHMSFPQ0YQ0QIILHQ00QPGYNNHHHHHHHHHNSOL 653
Qy 744 -----Q00000LSAQEALGKTRPKGYRNY-----DRDSLVEAVKAVOR 784

```







```

Db      ||| 361 RSLIDNRFKKMHQEQDHDODELSDSDAEAVDSNASTPVYPAPFAAQKRSHTL 420
Qy      ||| 421 SEHNSDGEVDYDSSPKKGRHPACGNASANOGAFASTPLDANVLLHMLAAGTAMPK 480
Db      ||| 421 SEHNSDGEVDYDSSPKKGRHPACGNASANOGAFASTPLDANVLLHMLAAGTAMPK 480
Qy      ||| 481 LDETQVDFPKGLLVANSAGGIMNGGLNLASASENGNGNLSLLQQOQHQQHQQHQQ 540
Db      ||| 481 LDETQVDFPKGLLVANSAGGIMNGGLNLASASENGNGNLSLLQQOQHQQHQQHQQ 540
Qy      ||| 541 QOQOQOQAAAYRHRLPKSETPETNSLDPNDASEDPILIKPSFKVSGPSSSSLPGLV 600
Db      ||| 541 QOQOQOQAAAYRHRLPKSETPETNSLDPNDASEDPILIKPSFKVSGPSSSSLPGLV 600
Qy      ||| 601 GGHHPPLANNNSLSISNSNSHSHRNGSNRSPHSASPMILAAVAQGYASGNSLTSS 660
Db      ||| 601 GGHHPPLANNNSLSISNSNSHSHRNGSNRSPHSASPMILAAVAQGYASGNSLTSS 660
Qy      ||| 661 SSSIOKMAASNTQRIINQSGESLRNGNVSDCSSNNGSSSLGKKPSISAKITIGTD 720
Db      ||| 661 SSSIOKMAASNTQRIINQSGESLRNGNVSDCSSNNGSSSLGKKPSISAKITIGTD 720
Qy      ||| 721 TSFSGASPNLSQOQHSAHHLTHQOQOQOQSAQOALGKTRPKRGKYNRYDRSL- 775
Db      ||| 721 TSFSGASPNLSQOQHSAHHLTHQOQOQOQSAQOALGKTRPKRGKYNRYDRSL- 775
Qy      ||| 776 -VEAVKAVQSGEMSHRAGSYGV-----PHSTLEKYKVERHMLRPKREKPKPDVLG 828
Db      ||| 776 -VEAVKAVQSGEMSHRAGSYGV-----PHSTLEKYKVERHMLRPKREKPKPDVLG 828
Qy      ||| 829 LT--GPAVKLQDLKAKGPH-----GSKYLSNALKNQNNQAAAAAATAATP 878
Db      ||| 829 LT--GPAVKLQDLKAKGPH-----GSKYLSNALKNQNNQAAAAAATAATP 878
Qy      ||| 839 QTEGTTWMLQAEQCPQEPKQSGGGGGSSSSGRCHAQRPETAPFPFGSGTGLSA-- 896
Db      ||| 839 QTEGTTWMLQAEQCPQEPKQSGGGGGSSSSGRCHAQRPETAPFPFGSGTGLSA-- 896
Qy      ||| 879 NGLKLPLEBAGFQALSFPQNMFWPQTNATNAYGLDNFRITTEAMRNPQASNNHG- 931
Db      ||| 879 NGLKLPLEBAGFQALSFPQNMFWPQTNATNAYGLDNFRITTEAMRNPQASNNHG- 931
Qy      ||| 932 -----LMSADQMVENVYDGIIRKTLQASSEGNSAAGNSGNSGNGHGHGHGHALLD 986
Db      ||| 932 -----LMSADQMVENVYDGIIRKTLQASSEGNSAAGNSGNSGNGHGHGHGHALLD 986
Qy      ||| 986 QLLVKTLPFTFNHNNNDYATCSASGESVYKSSGSPMGNYADIREALSADSGSSDDE 1046
Db      ||| 986 QLLVKTLPFTFNHNNNDYATCSASGESVYKSSGSPMGNYADIREALSADSGSSDDE 1046
Qy      ||| 1047 HSASHINNNSDLAANKKSGGGGGGGGNGGNGRSGSRMTSRDSETDASSFKSGENG 1106
Db      ||| 1047 HSASHINNNSDLAANKKSGGGGGGGGNGGNGRSGSRMTSRDSETDASSFKSGENG 1106
Qy      ||| 1046 HSASHINNNSDLAANKKSGGGGGGGGNGGNGRSGSRMTSRDSETDASSFKSGENG 1104
Db      ||| 1046 HSASHINNNSDLAANKKSGGGGGGGGNGGNGRSGSRMTSRDSETDASSFKSGENG 1104
Qy      ||| 1107 GQONHKMDLNGSSSS 1123
Db      ||| 1107 GQONHKMDLNGSSSS 1123
Qy      ||| 1105 GQONHKMDLNGSRAA 1121
Db      ||| 1105 GQONHKMDLNGSRAA 1121

```

## RESULT 2

```

A:3742 female sterile homeotic protein, 205k - fruit fly (Drosophila melanogaster)
N:Alternative names: membrane protein fsh, 205K
C:Species: Drosophila melanogaster
C:Date: 03-Mar-1993 #revision 03-Mar-1993 #text_change 20-Sep-1999
A:Accession: A43742; B43742
R:Haynes, S.R.; Mozer, B.A.; Bhatic-Dey, N.; Dawid, I.B.
Dev. Biol. 134, 246-257, 1989
A:Title: The Drosophila fsh locus, a maternal effect homeotic gene, encodes apparent men
A:Reference number: A43742; MUID:89276730; PMID:2567251
A:Accession: A43742
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2038 <HAY>
A:Cross-references: EMBL:M23221; NID:g157452; PIDN:AAA28540.1; PID:g157453

```

```

A:Accession: B43742
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1106 <HA2>
A:Cross-references: EMBL:M23222
C:Genetics:
A:Gene: fsh
A:Cross-references: FlyBase:FBgn0004656
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: alternative splicing; transmembrane protein
F:1-2038/Product: female sterile homeotic protein, 205K #status predicted <MA>
F:1-1106/Product: female sterile homeotic protein, 110K #status predicted <MA>
F:59-116/Domain: bromodomain homology <BRO1>
F:503-560/Domain: bromodomain homology <BRO2>

Query Match 5.28; Score 311.5; DB 2; Length 2038;
Best Local Similarity 21.1%; Pred. No. 9.7e-08;
Matches 236; Conservative 129; Mismatches 331; Indels 425; Gaps 50;

Qy      ||| 140 NAQKLVAESAAATSSSTSHIHQSDSNSASLPHHISSSSSNNNSGNRARIHA 199
Db      ||| 1064 SASKVEAVOPANVSSSSSSS-----DSSSSSD-----SSSDSDSEAGGDE--- 1110
Qy      ||| 200 AASARATPAATPANSLKYKLTQRAAKMTSMQMAOLAQFSLADFNLI--NSLASQ 257
Db      ||| 1111 -----RPPRKKSRSNNGSNVNPSTI-----NVVMGNLPS- 1141
Qy      ||| 258 QOQOQOQOQIASAVPTT-----SEVSAALISALKDTSPSPSDAPLDSLSSKPS-P 306
Db      ||| 1142 -----GALPTTLMGLDHVNVNSNTPTSGMSNMLGNANPLTAAAMNNNNKTSLP 1191
Qy      ||| 307 NSSISGVKSVACATPTPSGRAYSEEDLSRALQDY-VANKLDARASQHEORSILD 365
Db      ||| 1192 GSNFGG-----AAPFG-----NMHAAGVCPVAGAVASATGQHN----- 1227
Qy      ||| 366 NRLFQKHHDQDHDDELDSDNDAEAVDSNASTPVYPAPFA----- 410
Db      ||| 1228 -----KNGPNDLSKYQPGPPI-----NAALP--PHSPAGATATVATSSSGCI 1268
Qy      ||| 411 -RAQLKLSLSEHNSDLGED-----VDRGSPKKGRHPACGNASANOGAFAST-PL 460
Db      ||| 1269 RIASNLHFPQSL--GGDDLEHHAALAAALTSGINSTG--TAGGGINNNGSNNNANPL 1323
Qy      ||| 461 -----DANVLLHMLAAGTAMPKLDE--TQTVGDFIKGLLVANSAGGINNEGTL-NLS 512
Db      ||| 1324 GSGHGDAAVNASLASLSGLQIPQDPDVPQSLAS-----LEPSAGTGSGLTDNFI- 1377
Qy      ||| 513 ASQENSNGNASLTLQOQOQHQQH--HQOHHQOQOQOQHVAAVYRHRLPKSETPETNSLDP 569
Db      ||| 1378 -MQHLMQPAFPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1421
Qy      ||| 570 NDASEDPILIKPSFKVSGPSSSSSLSPGLVGGHHHPLANNNSISNSNH-----SSNS 625
Db      ||| 1422 -----DYVTEL-----LSKGAENVGNNG--NHLINFLDMAAAYVOQKPOOQOQA 1466
Qy      ||| 626 HRNGSNRSPHSASPMILAAVAQGYASGNSLTSSSSI-----QKMAASNTQRIIN 677
Db      ||| 1467 HNGGFN-----VADFGMAGFDGIAMTAASFLDLEBPLQOQOQOQOQOQOQH 1513
Qy      ||| 678 EQSGESLRNGNVSDCSSNNGSSSLGYKKPSISAKITIGTDSRFQASPNLSQOQHS 737
Db      ||| 1514 QOQOQOQTHQ-----VADFGMAGFDGIAMTAASFLDLEBPLQOQOQOQOQOQOQH 1528
Qy      ||| 738 AHHLTHQOQOQOQSAQOALGKTRPKRGKYNRYDRSLVEAVKAVQSGEMSHRAGSYG 797
Db      ||| 1529 QOH--HQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1548
Qy      ||| 798 VPHSTLEKYKVERHMLRPKRE-----PKPOPLVGTGPANQLDLK-LKAG 844
Db      ||| 1549 -----QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1600
Qy      ||| 845 PHGSKLSNALKNQNNQAAAAAATAATNPNGKLPLEAGPALSFPQNMFW-- 901
Db      ||| 845 PHGSKLSNALKNQNNQAAAAAATAATNPNGKLPLEAGPALSFPQNMFW-- 901

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on:           October 29, 2003, 12:25:26 : Search time 50 Seconds  
                       (without alignments)  
                       2240.730 Million cell updates/sec

Title:	US-10-016-768A-10
Perfect score:	6030
Sequence:	1 MHSSYEISLERVEECMGR.....ILHEKLAQIKAEQVDQADQL 1165

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

```
Database :
1: pir_76:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4839	80.2	1221	2	T13283	probable transcript
2	311.5	5.2	2038	2	A43742	female sterile hom
3	302	5.0	2248	2	A35938	profilaggrin - hum
4	299.5	5.0	2175	1	S03170	homeotic protein C
5	294	4.9	1182	2	T13804	shs protein - fruili
6	283.5	4.7	1180	2	S69305	stirbe a/b protein
7	282.5	4.7	1768	2	T13349	parallel sister ch
8	280.5	4.7	1741	2	T13510	parallel sister ch
9	279	4.6	2150	2	S71629	sensory transduct
10	275	4.6	1436	2	S57238	forked protein 5.4
11	275	4.6	1449	2	S57237	forked protein 5.6
12	271	4.5	1365	2	S14871	suppressor two of
13	270.5	4.5	1217	2	S52714	sericinB - silkw
14	269	4.5	2271	2	P90073	hypothetical prote
15	261	4.4	1085	2	S66149	Gene p19sqeak pro
16	259	4.3	1028	2	A56038	DNA-binding protei
17	257.5	4.3	1077	2	A44067	serine-rich protei
18	256.5	4.3	6713	2	B849921	hypothetical prote
19	249	4.1	1165	2	S62982	vacuolar protein V
20	249	4.1	1505	2	JC4851	hypoxia-inducible
21	248.5	4.1	3180	2	T13828	CREB-binding prote
22	248.5	4.1	1063	2	D86731	hypothetical prote
23	244.5	4.1	883	2	S04322	puff 74E protein -
24	244	4.0	1655	2	T13998	Gene mastermind pr
25	242.5	4.0	1385	2	T24276	hypothetical prote
26	242.5	4.0	4776	2	E95206	cell wall surface
27	238.5	4.0	1403	2	S24548	homeotic protein P
28	237	3.9	2248	1	D42088	adenylate cyclase
29	236	3.9	871	2	E53225	ecdysone-induced p

30	234	3.9	663	2	S21912	BRcore-01-z1 prote
31	234	3.9	2222	2	T13434	hypothetical prote
32	232.5	3.9	1457	2	T14577	protein kinase Yab
33	231.5	3.8	1388	2	T13741	hypothetical prote
34	231	3.8	1093	2	T13853	AF17 protein - hum
35	231	3.8	1394	2	B34598	edysome-induced p
36	231	3.8	2752	2	T20532	hypothetical prote
37	230	3.8	2578	2	A56922	transcription fact
38	229.5	3.8	864	2	A49070	edysome-inducibl
39	229	3.8	1668	2	T13748	sex comb protein -
40	228.5	3.8	2738	2	E88320	protein F07A11.6 f
41	228	3.8	1060	2	S33641	homeotic protein z
42	228	3.8	1213	2	S16356	ovo protein - fruit
43	226.5	3.8	1072	2	T13232	dechhund protein
44	226.5	3.8	1104	2	S59310	probable membrane
45	226	3.7	2559	2	A56923	transcription fact

## ALIGNMENTS

RESULT 1  
T13283  
probable transcription factor E93 - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 13-Aug-1999 #sequence 13-Aug-1999 #text\_change 17-Nov-2000

C;Accession: T13283

R; Baehrecke, E.H.; Thummel, C.S.

Dev. Biol. 171, 85-97, 1995

A/Title: The Drosophila E93 gene from the 93F early puff displays stage- and tissue-spec

A; Reference number: Z17648; MUID:96018744; PMID:7556910

A:Accession: T13283

A: Status: preliminary; translated from GB/EMBL/DBJ

A; Molecule type: mRNA

A;Residues: 1-1221 &lt;BAE&gt;

A: Experimental source: strain Canton S  
A: Cross-references: EMBL:U23686; NID:G888047; PID:G888048; FIDN:AA060220.1

A; Experimental source: SLDA in Canon 3

A:Gene: E93

A: Cross-ref

```

A:Map position: 3R

```

C;Function:

A;Description: probably acts in a stage-specific regulatory hierarchy in the salivary gl

Query	March	80.2%	Score	4839	DB	2	Length	1221
	Best Local Similarity	84.0%	Pred.	4	8e-234			
	Matches	972	Conservative	25	Mismatches	90	Indels	70
							Gaps	9
QY	1	MHISSEYISLERYAEBCMGROWKHYYODKLTCSHLNIEEQPIAIGSEDEPQYHNSK	60					
Db	1	MHISSEYISLERYAEBCMGROWKHYYODKLTCSHLNIEEQPIAIGSEDEPQYHNSK	60					
QY	61	EISQSNPHCKTENHRLLEQOHGSQLLEEDSENNOTSHDSSRTPTPGATSTPSPPEPI	120					
Db	61	EISQSNPHCKTENHRLLEQOHGSQLLEEDSENNOTSHDSSRTPTPGATSTPSPPEPI	120					
QY	121	DWPSAACNFCVNGRLITVYAOGKLVAESATATSSSTSSSHIHOHDSOANSASLPHH	180					
Db	121	DWPSAACNFCVNGRLITVYAOGKLVAESATATSSSTSSSHIHOHDSOANSASLPHH	180					
QY	181	SSSSSSNNNSSGGRRARHIAAASARATPPAATPANSLELYKLTORAAKMTSMDSMAOQA	240					
Db	181	SSSSSSNNNSSGGRRARHIAAASARATPPAATPANSLELYKLTORAAKMTSMDSMAOQA	240					
QY	241	QFSLLADFNILNSLASOOOQOOQOQIASAVTPTTSEVSAALISPALKOTPSPSVDAPLUL	300					
Db	241	QFSLLADFNILNSLASOOOQOOQOQIASAVTPTTSEVSAALISPALKOTPSPSVDAPLUL	300					
QY	301	SSKPSPNSSISGCVKSVACATPTPSGRARYSEEDLSRALODVAVANKLARKSASOHHO	360					
Db	301	SSKPSPNSSISGCVKSVACATPTPSGRARYSEEDLSRALODVAVANKLARKSASOHHO	360					
QY	361	RSJLDNLFQMKHDOEOHDHGDLEEDSDNDAAEAVDYNASTPVVYPAEFAAROLRKLSHL	420					

**THIS PAGE BLANK (USPTO)**

[illegible]

	A:Accession:	B43742
	A:Status:	Preliminary
	A:Molecule type:	mRNA
	A:Residues:	1-1106 <HA2>
	A:Cross-references:	EMBL: M21222
	C:Genetics:	
	A:Gene:	fsh
	A:Cross-references:	FlyBase: FBgn0004656
	C:Superfamily:	unassigned bromodomain proteins; bromodomain homology
	C:Keywords:	alternative splicing; transmembrane protein
	F:1-7038/Product:	female sterile homeotic protein, 205K #status predicted <MA2>
	F:1-1106/Product:	female sterile homeotic protein, 110K #status predicted <MA1>
	F:59-116/Domains:	bromodomain homology <BR01>
	F:503-560/Domains:	bromodomain homology <BR02>

  

Query Match	5.2%	Score 111.5	DB 2	Length 2038
Best Local Similarity	21.1%	Pred. No. 9.7e-08		
Matches 236	Conservative 129	Mismatches 311	Indels 425	Gaps 50

  

Oy	140	NAOGLVAESAATATSSSTNSHIOHSDSNSASLPNHIISSGSSNNNSGNRAHIA	199
Db	1064	SASKVEAVQANPVSSSSSSS-----DSSSSSSD-----SSSSSDSSDFACGDCE--	1110
Oy	200	AASARAIPAAATPANSLELYKLITQRAXKTSMDSMAQLAOFSLADFNLI--NLSAQ	257
Db	1111	-----RPRKKKSRRSNSNVNPPI---NVVMGSLPS-1141	
Oy	258	QQQQQQOOIAGAVTPPTT-----SEVSAALSPALKDPSPSVDAPLDISSKPS-P	306
Db	1142	-----GALSPTMLMDLHVNSNTPTSGMSMNLGANLTAAAMLNNNKTSIP	1191
Oy	307	NSGISGVKSVACATPTPGGRAVSEEDLSRALDY-VANKLDARKASAOHEORSILD	365
Db	1192	GSNEGG-----APAPG-----NMHAGAGVPVAGAASASTGOQH-----	1227
Oy	366	NRLPKMKHHDEOHDDLEDSDNDABEVDNASASTPVPAEFA-----	410
Db	1228	-----KNGPNDSLKKQPGPI-----NAALP--PHSFAAGTATVATSOGSGI	1268
Oy	411	--RAOLRKLSLSEHNSDLGED-----VDRCSPPKMRHPACSNASANOARASI-PL	460
Db	1269	RIASNLHKPSSL-----GGDLGEHNLAALALTSIGNSTG--TAGCGINNNGSNNNANPL	1323
Oy	461	----DANVLHTLMILAAGICAMEKLDE--TOTVGDFIKGLLVANSQGINNEGIL-NLLS	512
Db	1324	GGSHGDAAVNVASLASLGKIQIFOPDPVEQS LAS-----LEFSAGTGKSGLTDNFL-	1377
Oy	513	ASQENSNCNASLLIQOOHQH---HOQHNOOOOOOVAAVRRLPKSETPEFNSSLDP	569
Db	1378	-MOOHLMPARPQOOOQOOOQPFCHQOOOQOOOQOOOQOOOQH-----	1421
Oy	570	NDASEDPILKIPSKVSGPRASSSSLSLPGCLVGHNNHPLNNNSLSISNNSH-----SNS	625
Db	1422	----DVYTEL-----LSKGAEVNGVMG--NHLNFENLDMAALAAYOOKHPROOQOOQA	1466
Oy	626	HRSNSNRSPHSAPMLAAAVQCGYSAGNSLLTSSSGSI-----QKMASNIOROIIN	677
Db	1467	NNGFN-----VADFMAAGFRDGLNMTAAETLDERSLQOOQOMQOLOQHH	1513
Oy	678	EOSGOESLRNGNVSDCSSNNNGSSSLGKPKPSISVAKIIGSTDTRFGASPNLLSQOHS	737
Db	1514	QOOOQOOTH-----QOOOHO-----QOOOHO	1528
Oy	738	AHNLTHOOROOOLSAOEALGKSTRPKYKXRYNDRSLVEAVKAUVORGEMS VNRAGSYUG	797
Db	1529	QOH---HOQOOOQLTQOO-----LOOOO-----	1548
Oy	798	VPHSTEYKVVERHILMPRRKE-----PKRPORDVLGTGRANKLDLQK-LKAG	844
Db	1549	-----QQQQQQOHLQOOQOONHOANAKLLIPKPIESM--MSPRDQOLOQDHQVL	1600
Oy	845	PHGSKUSNALKNONODAAAAAAAATAATRGCLKLPFEFGROYLSTORYMMF---	901

**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (USPTO)**



[illegible]

Db 811 GFLPLPAFOFAAAQVAAAGCGRGCHYRFADSELOLPFGAASAGRLGESLLPKCDPRMEAKL 870  
QY 753 QEALGKGTGRPKCKGYRNYDRDSLVEAVKAVORGENSEVHRAGS-----YYGPHSTLEY- 805  
Db 871 OEML-----RYMMDKYANAQALDTL--HISRVRELLSTVHNIGQRLFAKYLIGLSOGTVSEL 924  
QY 806 -----KYKERILMRPKRE-----PKPRDL-----VGLTGPAN 834  
Db 925 LSKRPMDKLEKGRDSYRKHNAMACDDNAVMLKSLPKPKSGSLPQYAGRGAGGAGDD 984  
QY 835 KLQDLKLKAGPHGSGSKLSNALK-----NONNQAAAAAATAATPGLKPLF 886  
Db 985 SMSEDRI---AHILSEASLWKKSSVAQHREGRSRSHGSGEDSHSNESSKSPQOSTSPFF 1041  
QY 887 EAGQALSFQPNMFWPOTNATNAYGLDENRILTEAKRNPDASNNHGLMKSAQDVENVYDG 946  
Db 1042 KVENOLKQHO-----HLNPQAAAQOREREREOREQOOR--LRHDDQDMMARLYOE 1092  
QY 947 IIRKT-----LQASEGNGSAAGNSGNGCHGHGHGHALDOLL--VKKT 994  
Db 1093 LIATRPETAPRSLFLPSRLFGGAAGMPCASNA-----FRMAEDENMHVEREI 1143  
QY 995 LPFTNHRNDYAA-----TCSASGESYKRSQSPMGNTA 1028  
Db 1144 AKLQOHQOQOQAAQAOAPFNFSSLMALQOQVYLVNGAODVSLAAAKDKILNG----- 1196  
QY 1029 DIKERLSADSGSS--DEHSASHINNNSDLAHNKKSGSG-----GGCGGNG 1076  
Db 1197 --QRSSLEHSGSSCSKQGERDDAYPS--LHGRSSEGQTPAPRAPSPGRTGAG 1249  
QY 1077 -----QTNG-NRSSRMTSRDSETDASFEKSGENGQOHNKMDLNGSSSSSHIKC 1128  
Db 1250 APPTAAPRTGCASSNSAARPLSNSILRPALSSQGEFAATASPLRM-----ASITWSLI 1305  
QY 1129 ESEAAATGHSPGHHTTSLIHEKLAQIKAEVDQADOL 1165  
Db 1306 TQPPVTPHSTPQRPRTKAV--LPPITQOQFDMFNVL 1339

RESULT 5  
T13804  
shs protein - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 13-Aug-1959 #sequence\_revision 13-Aug-1999 #exec\_change 17-Nov-2000  
C:Accession: T13804  
R:Treisman, J.E.; Lai, Z.C.; Rubin, G.M.  
A:Title: Short-sighted acts in the decapentaplegic pathway in *Drosophila* eye development  
A:Reference number: 21767; PMID:96038094; PMID:755710  
A:Accession: T13804  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1212 <TRE>  
A:Cross-references: EMBL:L42512; NID:g833709; PID:g833710; PIDN:AAC41608.1  
C:Genetics:  
A:Gene: shs  
A:Cross-references: FlyBase:FBgn0010460

Query Match 4.9%; Score 294; DB 2; Length 1212;  
Best Local Similarity 20.2%; Pred. No. 3.8e-07;  
Matches 235; Conservative 139; Mismatches 421; Indels 368; Gaps 44;

QY 46 AGSEDEPSQVNHSSKEISQSNPNHCKTENHRLLEQHNQSQ--LLEEDSENNOTSHDSR 103  
Db 7 AASEDSGHOQOHOQO-----OOOHQOHOQOPLATTSTVAATSTVLANO 50  
QY 104 TPTPGATSTSPPE--PIDWRPSAKNFCVNGRLTVAAGKLVESAAT--ATSSSTS 159  
Db 51 SPTNSQASPSFNSQEAFLPLRQO-----SAAATVAAAAATVAATGTS 96  
QY 160 NSHIHQDSDSSASALPHHIISSSSNNSSGNR--ARHIAAASRAATPAATPANSLE 217  
Db 160 NSHIHQDSDSSASALPHHIISSSSNNSSGNR--ARHIAAASRAATPAATPANSLE 217

**THIS PAGE BLANK (USPTO)**

Db	97	000000HRSISINMFDRTVNAKFPKRASSNAGGNNPVRNNSMLTPRGVTIGTGCNIRK	156
OY	218	LYKLTORAAKKTSHNDMAAOLQTSYSLADEFULINSLASOQOQOQO-----OQTASAV	270
Db	157	LTKV-----SSLTSHNHFAVUCYPPSNIYONSNAGSNASALORTTSESLYLMMSRYAAGA	211
OY	271	TPTTSEVSAALISPYLKOTPRSPVDRPLDLSK-----PBPNSISGDVVS	316
Db	212	TPTT--VBRASSNSSLATSTSTSL-APKSSSSSGGNSTPPOOQOQOVLVSSNNSSSSNS	268
OY	317	VRACATPTPRGGRAYSEEDLSBALODVAVANKLDARKASQONHEORSILNDRPLFKMHNDQ	376
Db	269	FTKASSPNNGNARGSVGAATSAATGTAA-----AGSHNNQPHNNNNNNNNHNNH	321
OY	377	EQ-----	378
OY	322	000000TSLSGHNSLYVAGSASAGGGGGSGSSGTAAGTNRKPKTTSFETISVT	381
UY	379	-----DHGDLEEDSNDAEAV--DSNASTP-----VYPAEFARQ	413
Db	382	VONPKYLAAGOTGDESBADLDDESHDDNSRTIDLENETPSMEDTFSKEVEYUANAALST	441
OY	414	LRLKLSLSEHNG-----	431
Db	442	NARVPTSSOYGLVUVDPRIAPSLGQTIQNVQVNSDNIINNVGAVTPCGTKKKDDIKET	501
OY	432	VDRGSR-----KMGSR-----HACGASANOGAPASIPLDANVLHTLML	471
Db	502	QHRSERFVVKVKTSETEPFGRGKMMCMYDIDHSVGGNGNNEKTSSTSEAAAT--TDGG	560
OY	472	AAIGCA--MPKLDETQTVGDFIKGLYVANSOGIMNEGILNLLSASOENSGNASLYLOO	528
Db	561	AAGVAGSEAPAHKTTQS-----MLPPOKUNE--NHLEANSTDANNVAYAOQOQ	609
OY	529	QOHQOHQOH-----	549
Db	610	QOQOQOQOQOQITVGNALTKTLRVALYNVSHSSSVTRSRYNATVEFLSPNLLAQOQOQOQOOL-	668
OY	550	AYNRHLPKSETPETSSLDPNDASE-----	585
Db	669	-----FDSVNAALAASSPNRAGRPNNMDYARTAMOLQTLQOLKOREDAWDF--	716
OY	586	SGPASSSSLSPOG--LYGCHNHRPLNNNSLSISNNSNHSNSH--RNGSNRSPHSAPML	641
Db	717	PGAGGYAVYQNGDSAVGA-----ASNNSNAAAATGESQJSTSYVEEQOQOQOQPLSAPRLT	772
UY	642	A-----AAVAGGYSAGNSLITSSSSSIQKMMASNIQ---ROIWEGOGESLRNGVS	691
Db	773	POAARTFAAVNAG--QSPRFQLEOQOQOQOQOQATQIDGIVPQRFQOQOQOQTPQOSTA	830
OY	692	DSSSNNGSSSSLYGKPKPSISYAKIIGCTDTRSRFGASPNLISQONSHANHLTHQOQOQOOL-	750
Db	831	QOAAAANAATSAVTPARPOQTSNTSMAAVTTCOGOTMPLS-----HWTSTEOOQPNIG	884
OY	751	-SAOEALCKGTRPKRGKYRNYDRDSLVEAVKAVORGENSVHRAGSYUG-----VPHSTL	803
Db	885	AAAAAAAAGGTAA-----TSVAAPQAI P--TLQDSAPSTIDRQOLWVPOQOQ	931
OY	804	EYKVKERHLMR--RKREPKRODVLGLTGRPAKLOLDIKAKGRPHGSKSLNKLKONNOA	862
Db	932	QOQHOOEEOOQOPOOQOQPLPRNLIASASANNNNLTLNTV--VAATGEATNLTLTLEDA	990
OY	863	AAAAAAA--AAAAATPGLKTLPLFEAGRPQALSFQPNMFVQUTAA--TNAGLDFNRIT	918
Db	991	TAALAAAFATGAAAAATGATSAALATQOQIQLOOQOPNAESETBASGISAID-NKIE	104
OY	919	EMMRIPQASNNHGLKMSADNVE	941
Db	1050	QAM--DLVKSH--LMIIVAREVEE	1068

stripe a/b protein - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C:Date: 28-Oct-1996 #sequence #revision 13-Mar-1997 #ext\_change 21-Jul-2000  
 C:Accession: S69205  
 R:Frommer, G.; Vorbrueggen, G.; Pasca, G.; Jaeckle, H.; Volk, T.  
 EMBO J. 15, 1642-1649, 1996  
 A:Title: Epidermal egg-like zinc finger protein of *Drosophila* participates in myotube gu.  
 A:Reference number: S69205; MUID:96203082; PMID:861288  
 A:Accession: S69205  
 A:Status: preliminary; nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-1180 <PRO>  
 A:Cross-references: EMBL:U42403; NID:g1147788; PIDN:AAB02355.1; PID:g1147789  
 C:Keywords: alternative splicing

Query Match	4.74	Score 283.5	DB 2	length 1180
Best Local Similarity	20.64	Pred. No. 1,26	06	
Matches 236	Conservative 126	Mismatches 457	Indels 23	Gaps 41
QY	35	LNIEEQPIAIGSEDEPQYNNSSKEISQSNPNHCKTEHRLKEOHNQSGOLLEEDSEN	94	
DB	3	LTKRRQNEL-IVGSOQHPSATSSASSAGATSSPGLSGQMLNLVLA--GAGAAITNSTAN	58	
QY	95	NOTSHDSRTPTGATSTR---SPRRPIDMRPAKCNFCMNGLLTVNQGKVLASAA	151	
DB	59	SQDSLNTPTPLGLSRLNPLQAPARRPRRAVBRRAASGTFGGYOTAAARPLHNSRA	118	
QY	152	TATSSSTSNHNIHONDSNSASALPHNIISSSSSSNNSSGNRARNIAAASARATPAAT	211	
DB	119	TSEVSTARPPV---ELDEYVDLIQVQQLLDSSA-----AAAA	154	
QY	212	PANSLLEYKLLTQRAAMTSMDSMAAQLAQRSLIADRLNLINSLASQOQOQOQOIAAVT	271	
DB	155	AANNPRTTEOSVOOPQONTVVOQRPHQOQO-----OQOQOQOQOQVLLAKPR	201	
QY	272	PTTSEVSAALISPAALKQTPSPYDAPRLDLSKSP-----NSS	309	
DB	202	PRINLOKATEYAAQOLNAVESSSPGSRVLLDYPRLYKGNYNHNTPTRGEDVALMFGSN	261	
QY	310	ISGDVKSVRACATPTPSGRPAUYSEEDLSRALQDVYANKLDA--RKASQHNQEORSLDN	366	
DB	262	GTGGV-----AGTPSAAM-----IMEGLETTLVAPTHNAFLTLTAAAAHFNVLFSDT	309	
QY	367	RLFKMKNHDEQHDHGDGELSDNDDAEYVDSNASTVUYAEAFRAQDLRKLSHISENGS	426	
DB	310	CLFR-----STAAGSSGSPATYLSIPAQFG-----SHHGGA	343	
QY	427	DLIEDVDKGRPKMGRPACSNAMASQAPASIPLDYAVLNLH--TLMIAAGIGAMPKIDET	484	
DB	344	STGSSN-----SLNSSSTASSSTSSSLHNTYTASSAAAAAASAAAAAASAN	386	
QY	485	QTVGDFLKGLLVANSGLINNEG-----LMLLSAQENSQMASLLIQOQOHNQ	536	
DB	387	NSV-----LRANQIATPTQSGSPGHVAQVAPATASGSSASHLISLITSCQNSPTSS	440	
QY	537	QHNOQOQOQOQVAAVYHRLPKS-ETRETNSSLDPNDAEEDRLKIPFKVSGRASSLS	595	
DB	441	AVQDEVNKKOLIA---LRGLDILPTVTSSDLPFRFGPTTVVERPRTIGSIESEDI-SLE	496	
QY	596	P-----GGLVGNHNRPLNNNS-----LSI-----SNNNSHNSHNRGNSRPHSAPML	641	
DB	497	POVISVAAPLHSCSPKKEERSTRPLALIVKESSNNNSCMYUPOHNNNNNNNNNTTSSST	556	
QY	642	AAVAGGAGSAGNSLLTSSSSSIQKIMASIIQOQINBQSOEELRYCNVVDCCSNNCGSS	701	
DB	557	TTTSKO-----TTSESNTCEVGSRGNTQSHQOQO-QOQOLHNNTTSSNSMCHSH	605	
QY	702	SLGCKPRISVAKIIGTDTSRFGAR-----NLSSQO-----NHSANLTLH---	743	
DB	606	QOQOQOQO-----QOQOCHMSAPQOYQOQNLIQOQOQSGYNNNNNNNNNNNSQL	653	
QY	744	-----QOQOQOQOQOQALSGTSPKQKQKYNV-----DRQSLDAVAKAQR	784	

**THIS PAGE BLANK (USPTO)**

Db 654 OOLQQOQQOQQOQQOQQOQLHQQOQQOQHQQHQQHQAIGHQOOSQOQAKISYR 713  
QY 785 GEMSVHRRSGYGVPHSTLEYKVERHLMRPKREKRPQPDVLGLTG----PANKL---- 836  
714 G1FTT--TGNANMAAAAAAAQQOHHQQOQLPSPQLGVLAGPMSPPNSIGNSM 771  
QY 837 ---QLDKLKAQ-----HGSKLSNALKNQNNQAAAA-----AAAA 871  
Db 772 GLPSPDKTFQPFPLFSLPAHYATMQOQQOQQOQQOQQOQAAGAAPSPYDGRAAAA 831  
QY 872 AAAAATPNCILK-----PLFEAGP-----QALSPQNNF 900  
832 AAOHAELLLGLTMDCTPLLLKQPPSTAGSAGPFGGLDLSHSEQLQQOQVYRSPKIO 891  
QY 901 WPDNTATNAYGLDFNRIITEMRNPQ-----ASNHGKLKSAQDWENVYDGI 947  
892 WLDSPAD--YAQQOQQVQVQOQQOQQOQTTLVLPPTSSASSSNAALGVLLPKQENYPD-- 947  
QY 948 IRTKLOASEGNGSLAAGNGSNGNG-----HGHHGHGHALLDQLLYK 992  
948 ---MQPS--SNGTGYSGSGSGSSAAAAAAAVQLAEYSPSTSKGHEILSOVYQOS 1002  
QY 993 T-PL 995  
1003 TVPL 1006

RESULT 7

Parallel sister chromatids protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000

C:Accession: T13349

R:Gandhi, R.L.; Goldberg, M.L.  
submitted to the EMBL Data Library, November 1995

A:Reference number: Z17657

A:Accession: T13349

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1768 <GAN>

A:Cross-references: EMBL:U40214; NID:g1100982; PID:g1100983; PIDN:AAA91230.1

C:Genetics:

A:Gene: pasc

A:Cross-references: FlyBase:FBgn0004655

Query Match 4.7%; Score 282.5; DB 2; Length 1768;  
Best Local Similarity 19.4%; Pred. No. 2,3e-06;

Matches 238; Conservative 160; Mismatches 467; Indels 361; Gaps 45;

QY 48 GDEPPOVYHNSKEISQSNPNHCKTENHRLBOQHNSQLL-EEDSENNOTSHDSRTPT 106  
Db 203 ASQOPPEPSSASGQKQKPKR---EKKLKPEAPSPRVLGRARKAVNYREVEDERYPT 258  
QY 107 PGATSTSPPEPIDMRPAKCNFCVNGRLTLVVAQKLVAESAAATAT-SSSTNSHHQ 165  
Db 259 P-----TKDLIPKAGROPAAEVAAITTLAAASEATISS 292  
QY 166 HDSDSNSSASLPHHISSSSSNNNSGNRAHIAAASARATPAATPANSLELYKLTOR 225  
Db 293 TFGSPGSEPLRPPTASASASTSSQLPS---ASGSASNPSPASRTPPEHPVILRISKG 349  
QY 226 AAKTSMDSMAAQLQASLLADPLNLNSLASOQQOQQOQOQIAAVTPTTSEVSAALSPA 285  
Db 350 TSLRLVSTDS---EPPSSSPAHQNLQLSVTEEPEREGDETVPASTPKIT---VKPL 403  
QY 286 LKQTPSPSVAPADLUSKPPNSISGDVKSVRACATPTPSGRAAYSEEDLSRALQ---- 341  
Db 404 RPTPLAADVD-----GSSAAVGGASADSPFEERKSQGLERNEDEEEDDEEPEPEI 458  
QY 342 -----DVVANKLDARKSAQHNEQORSILDNRLFKMKHDOEQ 378  
Db 459 NYCTVKISIPKPKERLKLIKTDVIRNAIAKAAAAA---ESREKKSRSKKHG--KQL 513

QY 379 DHDGELESDNDAAEVDNSNASTP-----VYPAEFARAQLRKLSLSEHNGSDL 428  
Db 514 LAAGGAAAPASGATPAEINSEKTPSPHLALSANSOQAHTPSHLQHLQHPORGSAV 573  
QY 429 -----GEDVGRGPKRGNHRPACGNASANOGAPA----- 456  
Db 574 ISPTRSBHDFFDSOSVIGSISKGNSTPQLLAQAVQEDSCVIRSGSSVITSLETSON 633  
QY 457 ---SIPUDANVLLTML--AAGIGAMPKLDFTQVDFIKGLLVANSGGIMNEGLLN 509  
Db 634 SSLVAPPSDIESRLSESMIMMITIDAGTGAASAVETP-----LOEDILA 676  
QY 510 LLSASQENNGASLLLOOQOQHQQHQQOQQOQQOQOQAAVVRHRLPKSETPETNSLSP 569  
Db 677 VLRGEVPLRNGMTD-----PEPTEEDQOQPRATRGRRKA---NNNVV 720  
QY 570 NDASEDPILKITSFKVSGPSSSSSLPGVLGHHPLNNNSLSISNNSHNSHRNG 629  
Db 721 TEPATE--TRTRGAKGADATTAISP-----PTGRN---TRGTRGSKAEGEV 765  
QY 630 SNRSPHSAPMLAAVAQ-----GGYASGSLTSSSSSIQKMAASNTQRI 676  
Db 766 DMEVDETAMTTVPANEQLEQATLPPRRGRNAAARANNNNLASVNNINKIAAN-----L 820  
QY 677 NQSGQESLRNGVSDCSSNNGSSSLGYKXPSISVAXIIGGDTSRFGASPYLLSQOH 736  
Db 821 SAAEASRLAEGVA-----GGAASYGRKKNOQVTVL-----QOE 858  
QY 737 SAHLTHOQQOQOQSLAQEALGKTRPKRGKYRNYDRDSLVEAVKAVQGEMSVHRAGSYV 796  
Db 859 -----PVPEQETPRAE-----EQTPAK----- 878  
QY 797 GVPHSTLEYKVERHLMRPKREKRPQPDVLGLTGANKLQDKLKAQP-----HGSKL 851  
Db 879 -IPHTD-----HREHSPDHPDPDPDELNNNSNSSLQHDGSSSPPRDPFKKDK 929  
QY 852 SNALKQNQNNQAAAAAAATPNCGLK--LPFEAGPOLSFQPMFMPOTNATN 908  
Db 930 KRTLLDTCGAANAGAGAAAPRPSGBOKAVALVSKKGSIFKSALVPSQDAEQ 969  
QY 909 AYGLDENRIITEMARNQOASNHHGLMKSAQDWENVYDGIIRKTLQASEGNSAAGNSNG 968  
Db 990 A-----TVAKRH-----LYKHSWD-----AALEANG--GNTNSDA 1017  
QY 969 SNGNGHGHGHALLDQLLYKTPLPF--TNHRNDYATCSSASGCVKRSQSP-M 1024  
Db 1018 SNASAGGVVAGAKDHLNHLAAGKSDGDFDSSPSSNNNGSSSACSSA--STLRCDSPAL 1074  
QY 1025 GNYADIKRERLSADSG-----GSSDEHSHASHINNNSDLAHNKIK 1065  
Db 1075 G-----KISRLAKQGVPAITSTSSDAFDLDLEPIAGELDLERSAA-----GA 1116  
QY 1066 SCGGGGG--GNGQOTNGG--RSSRMTSRDSETDASFSKSGENGQONKMMDLNGCSS 1121  
Db 1117 SAGGTGATTGGGATGGGCVVRDKTK--DYVPVRNVKTA-----HOIQEIGBYOE 1167  
QY 1122 SSSHICSEEAATGHHSPGHHTTSTL 1147  
Db 1168 MDDDEVETIDALQPHNPATRCLSAL 1193

RESULT 8

Parallel sister chromatids protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000

C:Accession: T13610

R:Murphy, L.; Harris, D.; Barrell, B.

submitted to the EMBL Data Library, April 1999  
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A:Reference number: Z17668  
A:Accession: T13610  
A:Status: preliminary; translated from GB/EMBL/DBJ

**THIS PAGE BLANK (USPTO)**

A: Molecule type: DNA  
A: Residues: 1-1741 <MUR>  
A: Cross-references: EMBL:298269; NID:e1355202; PID:e1251076; PIDN:CAB10973.1  
C: Geneticals:  
A: Cross-references: FlyBase:FBgn0013432  
A: Introns: 348/3; 1219/3; 1500/3; 1557/2; 1587/1; 1650/3  
A: Note: EG-87B1.2

Query Match	4.7%	Score 280.5	DB 2	Length 1741
Best Local Similarity	19.1%	Pred. No. 2.8e-06		
Matches 238	Conservative 157	Mismatches 449	Indels 403	Gaps 46
QY	48	SEDESYQYNNHSSKETESQSNPNHCKTENRLEQNHGSGU--EEDSENNQTSHDSSRPT	106	
DB	203	ASEQEOQSSASGKQKQKPK--EKKLKEAPRPSRYLGRARKAVVYREVDEBERYPT	258	
QY	107	PGATSTPPEPPEPIDWRPSAKNFCVNRRLITLVNQGLVAESAAT--SSSTNSHHQ	165	
DB	259	P-----TKDLIPKAGQRAEVAATATLAAASSAFSS	292	
QY	166	HDSDNSASLIPHHISSSSSSNNNSGNRRARIIAASARATPAATPANSLELYKTOR	225	
DB	293	TFGSPGSESTLPPPTSPASASASTSQLPS--ASGASNPSPASRTPREPRIVLRISKG	349	
QY	226	AAKMTSMDSMAAQLAFSLDLDFNLINLAAQQQQQQQQQIAATPTTSEVSAATSPA	285	
DB	350	TSRLVSTDS--EEPPSSPHQNLQNLQNLSTVEEPARSGDETPAPATPKIT--VXPL	403	
QY	286	LKDPTRPSVDARPLDSSKPSPNSSISGVKSAACATPTPSGRARYSEEDISRALQ----	341	
DB	404	RPLTADSDVD-----GSSAAVAGASAGSFPERRQSLSLEPNDEEEDEEEDEEPEEI	458	
QY	342	-----DVANKLDARKASASQHNHEORSILNRLPKMKHDEQ	378	
DB	459	NYCTVKISPDKPKRERKLIIKTVDIRNAIKAAAA--ESRSKKRSKKHKH--KQL	513	
QY	379	DHODELEDSNDQAEVDSNASTP-----VPAEFAFARQLRKLSHSENGSGL	428	
DB	514	LAAAGCAAPGACATPAEINSEFKTTPSPYLALSEANSQQAHTPSHLQHLQHPORGSAV	573	
QY	429	-----GEDVDRGSPKXGRHPACGNASAGAPR-----	456	
DB	574	ISPTTRSDHDEDSQSVLGSISSKGNSTPQLLAQVQEDSCVIRSGSSVITSDLETQNH	633	
QY	457	-----SIPLDANVLHTLML--AAGIGAMPKLDFTQYGDFTKGLLVANSGIMNEGILN	509	
DB	634	SSLVAPSPDIESRLSEMMMTIDGATGAASAVETPR-----LOEDILTA	676	
QY	510	LLSASQESNGSNASLLLQQQQHQHQHQHQHQHQHQHVAAYRHR-----LPLKS	558	
DB	677	VLREVEVRLKNGNTD-----PEPTEEDQQQPKRATRGHGRKANNVDTVPAT	725	
QY	559	ET-----PEFNSSLDP-----NDASEDPILKTIPTFKVSG-PASSSSLS	595	
DB	726	ETRTGRGAKGADATTAATISPTGKRNTRGTRGSRKAEQVEMVEDEMMTVPAINEQLE	785	
QY	596	PGCLVGHHHHLLNNNSLSISNNNSNHSNHSNRGNSRPHASPLLAALVAQGISAGNS	655	
DB	786	QATL-----PRGRNRMAARANNNN-----	805	
QY	656	LITSSSSSSIQKMMASNIQRIQNEQSGESLRGNVSDSSNNNGSSSLGYKKPSIVAKI	715	
DB	806	LASVNNNNINKITAA--LSAAKASRLALEGVA-----CGAARSYGRGRKQOQYQV	854	
QY	716	IGCTDTSRFSGAPVLLSQOHHSAHLLTQOQQOQLSAQELGKGTGPKRGKYRYNDRDSL	775	
DB	855	L-----QQE-----PVPEQETPRDAEE--EQRPAAK-----	878	
QY	776	VEAVKAVQGESESVNRAGSYGVPRHSTLEVYKVERHLYRPRKREKPPRDLVGLTGPAHK	835	
DB	879	-----IHTTD-----HREHSDHDBDPPEDELNNNSNNS	908	
QY	836	LQDLKLKAGP-----HGGSKLSNALKQNNQAAAAAATATNGIK--LPLFE	887	

[illegible]

RESULT 9  
71629  
sensory transduction histidine kinase dhka - slime mold (Dictyostellium discoideum)  
C:Species: Dictyostellium discoideum  
C:Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 24-Sep-1998  
C:Accession: S71629  
R:Wang, N.; Shaulsky, G.; Escalante, R.; Loomis, W.F.  
EMBO J. 15, 3890-3898, 1996  
A:Title: A two-component histidine kinase gene that functions in Dictyostellium development  
A:Reference number: S71629; MUID:96524397; PMID:8670854  
A:Accession: S71629  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-2150 <MAN>  
A:Cross-references: EMBL:U42597  
A:Experimental source: strain AX4  
C:Genetics:  
A:Gene: dhka  
A:Map position: 6  
C:Superfamily: response regulator homologs  
C:Keywords: autophosphorylation; phosphoprotein; phosphotransferase; two-component regul.;  
F12027-21442/Domain: response regulator homolog <RRH>  
F12076/binding site: phosphate (Asp) (covalent) #status predicted

[illegible]



**THIS PAGE BLANK (USPTO)**

Df

292 SSTPTVNTCGAVNNNNNNNNNNSTGSLGAI PMDRSFDGINITTEESTGGNNSPPS 351  
| | | | |  
QY -----GSYYGV-----HSTLEYKKERHLRPRKRE--PKPODVLG-TGPA 833  
| : : : : :  
Db NCCSGNGSGMGIEPLSPRLNLSLNSGVNVSPRLNLHNLNSSSNLPPLSPRHINFIHINS 411  
| : : : : :  
QY 834 NKQLDLKLXAPRGSGKSLNALKNONQAATA-----AAAAAATAATPGLKLPFEA 868  
| : : : : :  
Db NLNNNNNNNNIIPNNPNNSNY---NSNNNYS PRANNHHNIS PRGSNI SPRSNGGSTII---- 465  
| : : : : :  
QY 889 GPOLSFQPMFPOTNATAYGLDFNRITEAMRNPOASNHGLMKSADOMVENYDGII 948  
| : : : : :  
Db 466 SPENISNNNNII I--NNINNNNILTPRRNSPRELVNPPTSPLATSLSTLPIVSU- 521  
| : : : : :  
QY 949 RKTLOASEGSGAAAGNSNGNGCHGHGHALLDOLLVTKTLPFTNHRNNDYAAT 1008  
| : : : : :  
Db 522 ----TSNNNNOSNNNTPSINNNGR-NGHCIOTISEITGNKFVVYVINGNNNNNT 575  
| : : : : :  
QY 1009 CSSASGESVVRKSCSPMGVADIKRERLS----- 1036  
| : : : : :  
Db 576 NNSTSNNNITTTNN 635  
| : : : : :  
Db 1037 -ADSGS-----SDPEHSAHTNNNSDLAHNKVSGGGGGGGGCGQTNGRSSRMTS 1089  
| | | | |  
Db 636 GAGSGCSLRNRKNDKDENDNGSNSNTTSSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 695  
| : : : : :  
QY 1090 RDSEETDASSFKSGENGQQQHMKMDLNGSSSSSHIKCESEAATGHHSPG 1140  
| : : : : :  
Db 696 NNN 744  
| : : : : :  
RESULT 10  
SS57238  
forked protein 5.4K - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C>Date: 10-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 22-Jun-1999  
C:Accession: S57238  
R:Hoover, K.K.; Chien, A.J.; Corcos, V.G.  
Genetics 135, 507-526, 1993  
A>Title: Effects of transposable elements on the expression of the forked gene of *Drosophila*  
A:Reference number: S57236; MUID:94063487; PMID:8244011  
A:Accession: S57238  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1436 <HOO>  
A:Cross-references: EMBL:X69871  
C:Genetics:  
A:Gene: FlyBase:f  
A:Cross-references: FlyBase:Fggn0000630  
'tions: 43/2; 104/3; 166/3; 564/3; 717/3; 825/1; 1001/2; 1121/3; 1204/3; 1274/2  
uperfamily: unassigned ankyrin repeat proteins; ankylrin repeat homology; EGF homology  
j-83/Domain: ankylrin repeat homology <ANI>

Query Match 4.6%; Score 275; DB 2; Length 1436;  
Best Local Similarity 19.2%; Pred. No. 4.2e-06;  
Matches 210; Conservative 146; Mismatches 369; Indels 368; Gaps 39;

QY 22 QWKHYODKLTCSH-----LNIEEQPIALAGEDEPDQYNHSKRISQ 64  
| : : : : :  
Db 136 RMWORSRKSLDKYKSPINDAAENOVBECLANLVHGTSVDYNGKSSORHKSOQULHQ 195  
| : : : : :  
QY 65 SNFNHCXTENRLEEOHNGSQLLEEEDSENNOTSHDSRTPT--PCATSTPSPPEPIDW 122  
| : : : : :  
Db 196 -----OHQQOQQOQQOQHLLSSSCNSNSNSKOTSRSNTIKSKSSLTSSDVEPFYL 246  
| : : : : :  
QY 123 RPQAKCNFCVNGURLTVNAOQKLVAESAAATATSSSTSNSHIHOHDSNSSASLPHHIS 182  
| : : : : :  
Db 247 HPALACGSGCGGGLCMGCMGCMGMGLGISRIKNSDA--LYSQOSSSEKXLVNGSS 303  
| : : : : :  
QY 183 SSSSSNNNSSGNRAHIAAARATPAAATPANSELVLYKLTLQRPAAKKTSMDSMAAOAQF 242  
| : : : : :  
Db 304 SLQGSSSSGGG-----GVGGGGSGGALLRPDGLVYVPM--RNGGYVNTSPNGSISGE 355  
| : : : : :

QY 243 SLI-----ADENULINSLASQOOOQOOQOIASAVTPTTSEVSAAAISPAKCTPPSVDAPL 298  
Db 356 SFFLHDDPDYHLQSGAGSGVRGLRLQOOOCEETAQO----- 388  
QY 299 DLSKSPSSNSSISGQVSKSVRACATPTPSGRAVYSEEDLSRALODVYANKLDARKASQHN 358  
Db 389 --ESRPPAATVESGQRVYHQADVTPA-----DGGSDSVSISASTRSSSSC 438  
QY 359 EQRSLIDNRLFKMKIHDOEOHDGDELEDSDNDDEAEVDSNASTPVYPAEFARQLRKL 418  
Db 439 AARVSI-----ATVAATAP-----G 455  
QY 419 HLEHNGSDLEDYDRGSPPKWGRHPRACGNASANAGAPASIPLDANVLTMLAAGIGAM 478  
Db 456 NIKNNN-----YKVTBASP-----AIRMAAGI--- 477  
QY 479 PKLDETQTVGDFIKGLLVANSNGIMNEGMLNLASQENSGNASLLLOOQOHOQH--QQ 537  
Db 478 -----TTT-----EDIVLYREESRKQHOQOOQOHOLOQOATQ 514  
QY 538 HHQOOOQOOHAAVYRRLPKSETPETNSSLDPRDASEDPILKIPFVKVSGPSSSSSPG 597  
Db 515 QOOOQOOOQOOLOHKYVNGRSRSDSGSH-----SSASASSTRST 555  
QY 598 GLV--GGHHPLNNNSLSTISNNSHSSSHRNGSRSPHSASPMI----- 641  
Db 556 DIVLOYSNMH--LNNKRMNNNNNSNNSSNIIAQSSSSSNNNNNSLNNRKSHSIIGLHS 613  
QY 642 --AAAVAGCGYSAGNSLLTS-----SSSIQKXMASNIQRQINEGSGOESLRCNVSD 692  
Db 614 SKYSSCLDNYSAKVNLKNOLLNGCIKSDTYESCPREDVAERTKOTHKSMIRNNLAD 673  
QY 693 CSNNNGSGSSSLGKPKPSIS-VAKIIGTDTSRFGASPNLLSQOHSANHLTHOOOQOLS 751  
Db 674 ASSNNNTSGSIN--NISNIGMNGGNOSR-----NLKRVSSAPPQWVL- 715  
QY 752 AQEALGCKTRPKRGKYRNYRDSLVLEAVKA--VQRGENSYHAGSGYGVPHSTLEYKVK 808  
Db 716 --AVVNNPPP-----PPLRPPLRAVQOQOQNNLSVDPQNS----- 748  
QY 809 ERHLMPKPKREKPPQPDVLGLTGPANKLOLDIKAKAPHGSSKXSNALKNONQAAAA-- 865  
Db 749 -----MTGPT--MHLQXNVFQPNQSP-----TNAAGTGAAPPP 780  
QY 866 --AAAAAATAATENGKLPLEAGPQALSFOPNMFWPOTNATNAGLDFNRITTEAMRN 923  
Db 781 VPAPRPVATEAVDSOSGLEV-----VEEPSLRPSLTVRGNNHRTWTISANKKAKLL-N 833  
QY 924 PQASNHGMLKSAQDMVENYVDGIIRKTLQASEGNSAA-----GNGSNGSNGNH 974  
Db 834 AGSTGSSIIAASSDDS-QSRYGGSVH-----AANGSAANGHFYGYSEGNKXQNSNANGN 886  
QY 975 GHGCHGHALDLQVLVKKTPLEFTNHRNNDYATCCSAGSESVYKSSSPMKGNVADIKER 1034  
Db 887 GNGNGGGG-----NYHSHPNQPHYAT-----GQPHQYTSLYG----- 919  
QY 1035 LSADGSGSSDEEH 1047  
Db 920 -----GGSSAEDEH 927

RESULT 11  
S57237  
forked protein 5.6k - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 10-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 22-Jun-1999  
C:Accession: S57237  
R:Hoover, K.K.; Chien, A.J.; Corcos, V.G.  
Genetics 135, 507-526, 1993  
A:Title: Effects of transposable elements on the expression of the forked gene of Drosophila  
A:Reference number: S57236; MUID:94063487; PMID:8244011  
A:Accession: S57237

**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (USPTO)**

Db	433	ETYAKNIIGLKPIEQLOQAS-NPDKSYSPASMSSCSSSTNNSSSLGTADASTS---	490
Qy	343	VANKLDARKASQHEQRSILDNRLFKMKHDOEQDHDGELE-----DSNDAAEAVD	397
Db	491	-----TSTSSSHRKR-----KKKSKSRPKANGCRKKLHAETISSOTGKMVK	533
Qy	398	SNASTVYPAEFARA-----QRLKLSHEINGSDLEGVDVGRGKMKRHPACGAS	449
Db	534	ITAK-PNHKLDPRSHSLASGELDLOKL---KLDSTSTSEALNR---TLDEARSTISL	585
Qy	450	ANOGAPASIPLDANVLLHTLMLAAGIAMPEKLDGTQTVGDFIKGLVANSGIMNELLN	509
Db	586	VVGATTPPTPT-----AEPQQQQQQ-----	608
Qy	510	LISASQENSGNASLLQQQOHOQHQQOQQOQOQOQVAAVYRRLPKSE---TEETNSL	567
Db	609	-----QQOQQPQQOQQOQQOQQOQQOQFVV---LPKIKDLTPTSPPL	646
Qy	568	DPVDAEDPILKTPSFVSGPASSSLSPGLVGCH-----	603
Db	647	PPS-----LFKAYTP-STTPAHPHTVAGGKPKQOQOQMPQOPQAVLQOSMAKTN	694
Qy	604	--HHPLNNNSLSISNNSN-----SSNHRGSRSPHSPAMLAAVAGYAGNLL	657
Db	695	PAKRPPLSNNNR--KPNSGHFAVQPAQTHRRMYMOKYQSTPSSIA-----SAANKP	745
Qy	658	TSSSS-----SIQMMASNIOQDINSGOES-----LRGNVSDC	693
Db	746	KRSMSEDHPAKQARLSQAQAMASVAAKLHMQTSNOAKQAALFLPNQMRSGYLPDL	805
Qy	694	SSNNGSSSLGYKKPSTSVAKIIGTQTSRFGASPNLLSQOHNBAHLTHQOQQOQLSAQ	753
Db	806	---GSKPTLPMLCPASSSQOV---TTPPRRAPTSYSEEPNIHPALETVLRPVNQ	858
Qy	754	EALGKG-TREPCKGYRNYDRDLSLEAVKAVORCEMSYHARASYGVF-HSTLEKVKERH	811
Db	859	SAGCKGLTMPRLS-----PRATSSARLMGPRALRPHAGHGAARAS	901
Qy	812	LMRPKREKRP-----OPDLVGLTGPANKQLDQLKAGPHGSKLNL-	853
Db	902	COMPTMPRLPLRPMPTTIPAIKVSPPLSVALSGORNN-----KGNSSNNAV	951
Qy	854	-----LKKONNAAAAAATAATATPGLYLPFGAGROLSPQNMFWQNTN	908
Db	952	RTSPPALINLRNTAAPHSFPKSSPKVEASKSPR-ALCCQG-----KTGTA	1000
Qy	909	AVGLDENRILTEARMNPQASNNHGHGMSKAODVEVUYDGIIRKTLQASEGNSAANGSNG	968
Db	1001	A--LDSKTS-----LREFRPAVQAV-----TATATTSVTTAAGAG--	1036
Qy	969	SNNGHGHGHGHCHAL---LDQLLVKTPPLFTNHRNNDVATCSSAGSVKSGSPMG	1035
Db	1037	-----GAGTGTGTALAKDADIDLSANP-----GRSNN	1076
Qy	1026	NYADIKRERLSAOSGSSDEHSHIINNNS---DLANKK-----SGG-----G	1070
Db	1077	N-----NNNNNNNNNNNNNNNNNNNNSTNSLEAALNKIKONISANSNGPSTTS	1125
Qy	1071	GGGNGGOTNGG--RSSRMTSRDSEDTASSFGKSGENGQOQNHKMDLNG--GSSSSHIK	1128
Db	1126	GSNSNGTNDGDLQNLHMLSEBATAREKISTKAASG-----NGSSTSSSKPR	1175
Qy	1129	ES-----EATGHHSPGHHTTSLIHEKLAQIKAEQ	1158
Db	1176	KANALVPRPONASVRSIPNPSALAFRNPQAPAAASTASISKULTYRAEE	1223

**THIS PAGE BLANK (USPTO)**





**THIS PAGE BLANK (USPTO)**

[illegible]

Search completed: October 29, 2003, 12:30:34  
Job time : 140 secs

**THIS PAGE BLANK (USPTO)**